Homotopy Methods for Linear Optimization Problems with Sparsity Penalty and Applications

Von der
Carl-Friedrich-Gauß-Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von
Christoph Brauer
geboren am 1.3.1985
in Stade

Eingereicht am: 23.11.2017
Disputation am: 19.03.2018
1. Referent: Prof. Dr. Dirk Lorenz
2. Referent: Prof. Dr. Marc Pfetsch

2018

Zusammenfassung

Gegenstand der vorliegenden Dissertation ist eine primal-duale Homotopiemethode für das konvexe Optimierungsproblem $\min \|\boldsymbol{x}\|_1$ s.t.$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \delta$. Darin können wir $\boldsymbol{b} \in \mathbb{R}^m$ als eine Messung interpretieren, welche über einen durch die Matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ beschriebenen linearen Messprozess aus einer nicht beobachtbaren Größe $\boldsymbol{x} \in \mathbb{R}^n$ hervorgeht. Ziel des oben genannten Optimierungsproblems ist die Rekonstruktion des unbekannten Vektors $\boldsymbol{x}$.

Die Nebenbedingung im oben genannten Optimierungsproblem ist motiviert durch die Annahme, dass bei der Messung ein additiver Fehler $\boldsymbol{\eta}$ entsteht, sodass $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{\eta}$ sowie $\|\boldsymbol{\eta}\|_\infty \leq \delta$ gelten. Ist die Anzahl der Messungen m kleiner als die Dimension n der Zielgröße und hat die Matrix $\boldsymbol{A}$ vollen Rang, so ist bereits das Gleichungssystem $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ unterbestimmt und besitzt unendlich viele Lösungen. Gleiches gilt folglich für das zur Nebenbedingung äquivalente System linearer Ungleichungen $-\delta\mathbb{1} \leq \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \delta\mathbb{1}$. Ohne zusätzliche Informationen wäre es demnach unmöglich, die gesuchte Größe $\boldsymbol{x}$ verlässlich zu schätzen. Die Zielfunktion erklärt sich nun durch die zusätzliche Annahme, dass $\boldsymbol{x}$ dünn besetzt ist, also nur wenige von Null verschiedene Einträge besitzt. Eine zentrale Erkenntnis aus dem Themengebiet *Compressed Sensing* besagt, dass die Lösung mit minimaler ℓ_1-Norm, unter gewissen Voraussetzungen, gleichzeitig die Lösung mit den wenigsten von Null verschiedenen Einträgen ist.

Im ersten Teil dieser Arbeit nutzen wir zunächst Techniken aus der konvexen Optimierung, um primal-duale Optimalitätsbedingungen für das oben genannte Problem herzuleiten. Darauf aufbauend, motivieren wir ein iteratives Verfahren, welches durch sukzessives Verkleinern des Homotopieparameters δ eine optimale Lösung des Optimierungsproblems liefert. Dabei nutzen wir aus, dass für den Startwert $\delta = \|\boldsymbol{b}\|_\infty$ die eindeutige Lösung durch $\boldsymbol{x} = \boldsymbol{0}$ gegeben ist und verkleinern δ anschließend derart, dass simultan stets eine zugehörige optimale Lösung $\boldsymbol{x}$ berechnet werden kann. Der Begriff der *Homotopie*methode bezieht sich darauf, dass der so entstehende Lösungspfad stetig und stückweise linear ist. Insbesondere liefert das beschriebene Verfahren den gesamten Lösungspfad für das Intervall $[\delta, \infty)$. Weiterhin zeigen wir, dass unser Verfahren stets nach einer endlichen Anzahl von Schritten terminiert und dabei eine optimale Lösung liefert. Anschließend beweisen wir, dass einerseits die Anzahl der benötigten Iterationen nach oben durch $(3^{m+n} + 1)/2$ beschränkt ist, und andererseits tatsächlich Instanzen existieren, für deren Lösung der Algorithmus exakt $(3^n + 1)/2$ Iterationen benötigt.

Im Anschluss an die Analyse unserer Methode diskutieren wir die Äquivalenz des oben genannten Optimierungsproblems zu einem bestimmten linearen Programm und stellen eine Erweiterung unseres Verfahrens für ℓ_1-Norm Minimierungsprobleme mit beliebigen linearen Nebenbedingungen vor. Insbesondere folgt dann, dass unsere Methode auch als Löser für lineare Programme mit echt positiven Zielfunktionskoeffizienten geeignet ist. Schließlich stellen wir verschiedene Anwendungen aus den Themengebieten Signalverarbeitung, Maschinelles Lernen und Statistik vor, an welchen wir die Effektivität und die Effizienz unserer Methode mittels numerischer Beispiele demonstrieren.

Contents

1 Introduction **7**

2 Preliminaries **11**
2.1 Notation 11
2.2 Duality in Convex Optimization 12

3 Homotopy Method **17**
3.1 Optimality Conditions 17
3.2 Homotopy Approach 18
3.2.1 Dual Updates 18
3.2.2 Primal Updates 19
3.3 A Theorem of the Alternative 20
3.4 ℓ_1-HOUDINI Algorithm and Finite Termination 25
3.5 Analysis of the Solution Path 27
3.6 Upper Complexity Bounds 34
3.7 Lower Complexity Bounds 34

4 Active-Set Methods **45**
4.1 Active-Set Method for Linear Programs 46
4.1.1 Optimality Conditions 46
4.1.2 General Theme 47
4.1.3 Descent Directions and Blocking Constraints 48
4.1.4 Lagrange Multipliers 48
4.1.5 Feasibility of Generated Directions 49
4.1.6 Algorithm and Set Management 49
4.2 Active-Set Method for the Dual Update 51
4.2.1 Initialization 51
4.2.2 Descent Direction and Blocking Constraints 52
4.2.3 Lagrange Multipliers 53
4.3 Active-Set Method for the Primal Update 53
4.3.1 Initialization 53
4.3.2 Descent Direction and Blocking Constraints 54
4.3.3 Lagrange Multipliers 55
4.4 The Ambiguity of Lagrange Multipliers 55
4.4.1 Initial Direction for the Dual Update 56
4.4.2 Initial Direction for the Primal Update 57

5 Connection to Linear Programming and Extensions **59**
5.1 Associated Linear Programs 59
5.2 Parametric Simplex Method 60
5.3 Extension to Non-Uniform Constraints 62
5.3.1 Two-Sided Inequality Constraints 62
5.3.2 Arbitrary Linear Constraints 62

6 Applications **69**
6.1 Speech Coding 69
6.1.1 Encoding 70
6.1.2 Decoding 71
6.1.3 Uniform Quantization 72
6.1.4 Non-uniform quantization 73
6.1.5 Numerical Experiments 75
6.1.6 MAP Estimation 78
6.2 The Dantzig Selector 83
6.2.1 Optimization Problem and Motivation 84
6.2.2 Algorithmic Approaches 84
6.2.3 Numerical Experiments 85
6.3 Sparse Precision Matrix Estimation 86
6.4 Sparse Linear Discriminant Analysis 87
6.5 Model Selection 88
6.5.1 General Cross-Validation Scheme 89
6.5.2 Grid Independent Cross-Validation 89
6.5.3 Numerical Experiments 92
6.6 The L1-Testset 93

7 Conclusion **99**

1 Introduction

In many recent applications, for instance in statistics and technology, the challenge arises to infer certain quantities of interest from measured information. If the measurement process can be modeled in terms of a linear mapping $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then the associated problem is to find a solution of the linear equation system $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{b} \in \mathbb{R}^m$ is a measurement vector and $\boldsymbol{x} \in \mathbb{R}^n$ is the sought-after quantity. It is well-known that, in case $m < n$ and the matrix has full rank, the above-mentioned linear equation system has infinitely many solutions. This makes the search for $\boldsymbol{x}$ infeasible as long as no further information is available.

A fundamental concept in the field of *Compressed Sensing* [10, 16, 18, 20] which can make the search for $\boldsymbol{x}$ feasible is *sparsity*. A signal is called sparse in case it has only few non-zero components, and it can be observed that many real-world quantities are indeed sparse or sparsely approximable (see, e.g., [20]). In case $\boldsymbol{x}$ is not sparse itself, it may still be that it has a sparse representation in terms of some known *dictionary* $\boldsymbol{D} \in \mathbb{R}^{n \times k}$, i.e., there exists a sparse coefficient vector $\boldsymbol{a}$ such that $\boldsymbol{x} = \boldsymbol{Da}$. With this knowledge at hand, it is natural to search for the sparsest solution of the linear equation systems $\boldsymbol{Ax} = \boldsymbol{b}$ or $\boldsymbol{ADa} = \boldsymbol{b}$, respectively. In the following, we assume for simplicity that the first case applies, i.e., we seek for a quantity $\boldsymbol{x}$ which is assumed to be sparse itself.

The sparsest solution of a linear equation system can be determined in terms of the combinatorial optimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \|\boldsymbol{x}\|_0 \quad \text{s.t. } \boldsymbol{Ax} = \boldsymbol{b} \tag{SP}$$

which is known to be NP-hard (see [33] among others). With the advent of compressed sensing, the recovery of sparse vectors by means of the *Basis Pursuit* approach [13]

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \boldsymbol{Ax} = \boldsymbol{b} \tag{BP}$$

became popular. Basis Pursuit can be interpreted as the convex relaxation of (SP) (see [20]) which is why there exist a variety of efficient algorithms to solve (BP).

However, in many applications, measurements are degraded by some kind of *additive noise* $\boldsymbol{\eta} = \boldsymbol{Ax} - \boldsymbol{b} \in \mathbb{R}^m$ which needs to be taken into account. In that context, one very popular approach is the so-called *Basis Pursuit Denoising* (or ℓ_1*-regularized Least-Squares*) problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \lambda \|\boldsymbol{x}\|_1 + \tfrac{1}{2} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2, \tag{ℓ_1-LS}$$

with $\lambda > 0$, which received a lot of attention over the past decade (see, e.g., [28, 20] and many references therein). Note that, in a slightly different formulation, (ℓ_1-LS) is also known as the *Least Absolute Shrinkage and Selection Operator* (LASSO) [42]. However, the related problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_\infty \leq \delta \tag{P_δ}$$

appears to be much less investigated, both theoretically and algorithmically. This problem can be rewritten as a linear program (LP) by formulating the ℓ_∞-norm constraint as linear inequalities and performing the usual variable split of $\boldsymbol{x}$ into its positive and negative parts (cf. Section 5.1). Thus, in principle, every LP solver can be applied to solve the problem.

However, in practice, it may happen that the problem instances are very large (and with $\boldsymbol{A}$ dense or perhaps only available implicitly) so that current LP solvers may not be able to handle the problem well. Moreover, there are cases in which one does not only want to solve the problem for a given instance of $(\boldsymbol{A}, \boldsymbol{b}, \delta)$ but for a whole range of parameters δ. In this thesis, we propose a new homotopy algorithm for the problem (P_δ) which we call ℓ_1-HOUDINI (HOmotopy UnDer Infinity Norm ConstraInts). Our method exploits the fact that the solution path with respect to the homotopy parameter δ of the problem (P_δ) is continuous piecewise linear (cf. Section 3.5). As a consequence, the solutions associated with all parameters $\delta \geq 0$ for which the feasible set of (P_δ) is non-empty can be calculated by performing only one call of our algorithm.

Homotopy concepts have been around for decades, so it should come as no surprise that our approach bears some resemblance to several earlier algorithms. There exists a variety of homotopy schemes for LPs (see, for instance, [3, 34] and references therein). In fact, the latter work shows how many standard LP algorithms (simplex, affine-scaling and interior-point methods) can be subsumed under a unifying homotopy framework, exhibiting nice connections between intuitively very different approaches. A specific homotopy scheme for the LP reformulation of (P_δ) is discussed in [44] and [37] (cf. Section 5.2). Moreover, a homotopy algorithm for (ℓ_1-LS) has been proposed in [36] and an approach for the *Dantzig selector* problem (see below) has been introduced in [1].

The remainder of this thesis is structured as follows: In Chapter 2, we start by fixing some notation and then recapitulate those basic ideas from convex optimization which are most relevant for our approach. In particular, we use the well-known concept of *Fenchel-Rockafellar duality* in order to derive primal-dual optimality conditions for the problem (P_δ).

Proceeding from these conditions, we introduce our idea of a dedicated homotopy scheme for (P_δ) in Chapter 3. Afterwards, by means of a theorem of the alternative, we show that our ℓ_1-HOUDINI algorithm terminates after a finite number of iterations yielding an optimal solution of (P_δ). The derivation of our algorithm as well as the proof of finite termination originate in [5]. In this thesis, we extend our analysis of (P_δ) to the solution path. On the one hand, we show that the number of linear segments in the path, which is equal to the number of iterations that ℓ_1-HOUDINI needs to perform in order to find an optimal solution, is bounded above by $(3^{m+n}+1)/2$. After that, we

adapt the worst-case analysis for the LASSO that has been introduced in [31], and show that, for each $n \in \mathbb{N}$, there exist instances of (P_δ) where the solution path has exactly $(3^n + 1)/2$ linear segments.

It will turn out that ℓ_1-HOUDINI essentially consists of alternating primal and dual update steps in the form of specific linear programs. As a consequence, our algorithm is easy to implement as long as one has access to an arbitrary LP solver. However, in Chapter 4, we make use of the observation that the LPs occuring in the primal and dual updates of ℓ_1-HOUDINI have a certain structure, and develop a dedicated active-set algorithm for linear programming. Using this active-set algorithm turns out to be particularly efficient because it allows, in some sense, that useful information about *improvement directions* is passed between consecutive primal and dual update steps. Note that the presented active-set method is part of [5] as well.

At the beginning of Chapter 5, we show in detail that (P_δ) and the associated dual optimization problem have equivalent LP reformulations, and discuss the homotopy method introduced in [44, 37] which is based on the LP reformulation of (P_δ). Afterwards, we develop an extension of ℓ_1-HOUDINI which enables the algorithm to treat ℓ_1-norm minimization problems with arbitrary linear constraints. We conclude the chapter with the statement that the extended version of our method can also treat arbitrary LPs as long as these have throughout strictly positive objective function coefficients.

Our interest in sparse approximation under ℓ_∞-constraints is motivated by several practical applications, some of which we present in Chapter 6, complemented by diverse numerical experiments and demonstrations: The Dantzig selector problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\|_\infty \leq \delta \tag{DS_δ}$$

is a special case of (P_δ) and has numerous applications in statistical estimation, see, e.g., [46], where the whole solution path for $\delta > 0$ is computed as a selection step prior to a classification step. In *sparse dequantization*, one has quantized measurements $\boldsymbol{b} = Q(\boldsymbol{A}\bar{\boldsymbol{x}})$ of some signal vector $\bar{\boldsymbol{x}}$ which is assumed to be sparse. If the quantization level is known, one can interpret (P_δ) as the problem of finding a reconstruction $\boldsymbol{x}^\star$ with minimal ℓ_1-norm for which the measurements $\boldsymbol{A}\boldsymbol{x}^\star$ produce the same quantized measurements $\boldsymbol{b}$. We refer to [24] for the general idea and to [4] for a recent application to speech processing (with the involvement of the author). In *sparse linear discriminant analysis* as proposed in [7], one obtains a problem of the form (P_δ) in which $\boldsymbol{A}$ is a sample covariance matrix and $\boldsymbol{b}$ is a difference of samples means. Similarly, the so-called CLIME estimator [8] solves *sparse precision matrix estimation* problems via a sequence of (P_δ) problems in each of which $\boldsymbol{A}$ is again a covariance matrix and $\boldsymbol{b}$ is equal to a unit vector.

Chapter 7 finally concludes this thesis.

2 Preliminaries

2.1 Notation

For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the vector $\boldsymbol{a}_i^\top$ denotes the i-th row and $\boldsymbol{A}_j$ denotes the j-th column. Moreover, for $I \subseteq \{1, \dots, m\}$ and $J \subseteq \{1, \dots, n\}$, $\boldsymbol{A}_J^I$ denotes the sub-matrix with rows indicated by I and columns indicated by J. With respect to the transposed matrix, we sometimes write $\boldsymbol{A}_J^\top = (\boldsymbol{A}_J)^\top$. In case of a vector $\boldsymbol{x} \in \mathbb{R}^n$, the j-th entry is denoted by x_j. Matrices and vectors are throughout printed in bold, while single numbers and sets are printed normally. By the symbol $\odot$, we denote the component-wise product of two vectors $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$, i.e., $(\boldsymbol{x} \odot \boldsymbol{z})_j = x_j z_j$, and $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \boldsymbol{x}^\top \boldsymbol{z}$ refers to the inner product of the vectors. Furthermore, we define $\mathrm{Diag}(\boldsymbol{x})$ to be the $n \times n$ diagonal matrix having the entries of the vector $\boldsymbol{x}$ as its diagonal elements.

As usual, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the respective norms, i.e.,

$$\|\boldsymbol{x}\|_1 = \sum_{j=1}^{n} |x_j| \quad \text{and} \quad \|\boldsymbol{x}\|_\infty = \max_{j=1,\dots,n} |x_j|. \tag{2.1}$$

While $|x|$ naturally stands for the absolute value of a real number x, we write $|\boldsymbol{x}|$ to refer to the component-wise absolute value of a vector $\boldsymbol{x}$, i.e., $|\boldsymbol{x}|_j = |x_j|$. For a convex set $C \subseteq \mathbb{R}^m$, the associated *indicator function* is defined as

$$I_C(\boldsymbol{y}) \coloneqq \begin{cases} 0, & \boldsymbol{y} \in C \\ \infty, & \text{else} \end{cases}. \tag{2.2}$$

In case $C = \{\boldsymbol{y} \in \mathbb{R}^m : g(\boldsymbol{y}) \leq \alpha\}$ is a level set associated with some convex function g and some fixed $\alpha \in \mathbb{R}$, then we also refer to the related indicator function as $I_{g \leq \alpha}$. Slightly different, the *characteristic function* of a convex set is defined as

$$\mathbf{1}_C(\boldsymbol{y}) \coloneqq \begin{cases} 1, & \boldsymbol{y} \in C \\ 0, & \text{else} \end{cases}. \tag{2.3}$$

As in case of the indicator function, the term $\mathbf{1}_{g \leq \alpha}$ is sometimes used to indicate that C is a level set. The indicator function I_C has its values in the extended real numbers $\mathbb{R}_\infty \coloneqq \mathbb{R} \cup \{\infty\}$. In addition, we write $\mathbb{R}_\pm \coloneqq \{x \in \mathbb{R} : x \gtrless 0\}$ and $\mathbb{R}_\pm^0 \coloneqq \mathbb{R}_\pm \cup \{0\}$.

The all-zero vector in $\mathbb{R}^n$ is denoted by $\mathbf{0}$, and $\mathbb{1}$ stands for the n-dimensional vector containing only ones. Moreover, $\boldsymbol{I}_n$ refers to the $n \times n$ identity matrix whose colmuns are denoted by $\boldsymbol{e}_j$.

Further notation are introduced occasionally in the respective sections of this thesis.

2.2 Duality in Convex Optimization

A major part of this thesis addresses the minimization of a certain non-smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set of points $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $\boldsymbol{A}\boldsymbol{x} \in C$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a linear mapping and $C \subseteq \mathbb{R}^m$ is a convex set. Employing the *indicator function* $I_C(\boldsymbol{y})$ which is convex as well, we can reformulate the problem of minimizing f over $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \in C\}$ as the unconstrained problem of minimizing the sum $f(\boldsymbol{x}) + I_C(\boldsymbol{A}\boldsymbol{x})$. This kind of problem arises in numerous fields and applications (some of which will be discussed subsequently), and there is a dedicated and rich theory that we can draw on. In the following, we gather some results which are fundamental for our work. Theorem 2 states that under certain assumptions, the *primal problem* of minimizing $f + g \circ \boldsymbol{A}$ is associated with a *dual problem* which has the same optimal value. Afterwards, Theorem 5 provides sufficient and necessary *primal-dual optimality conditions*. We refer to [38] for a more detailed discussion of the subject and especially for the proofs of both theorems.

In preparation for the following theorems, we define the *domain* of $f : \mathbb{R}^n \to \mathbb{R}_\infty$ as $\operatorname{dom} f \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < \infty\}$ and call f *proper* in case $\operatorname{dom} f \neq \emptyset$. Further, f is said to be *lower semi-continuous at* $\boldsymbol{x}$ if and only if

$$f(\boldsymbol{x}) \leq \liminf_{n \to \infty} f(\boldsymbol{x}_n) \tag{2.4}$$

holds for each sequence with $\boldsymbol{x}_n \to \boldsymbol{x}$. Accordingly, f is called *lower semi-continuous* if it is lower semi-continuous at each $\boldsymbol{x} \in \mathbb{R}^n$. Finally, the *relative interior* of a set $C \subseteq \mathbb{R}^n$ is defined as

$$\operatorname{ri} C \coloneqq \{\boldsymbol{x} \in C \mid \exists\, \varepsilon > 0 : B_\varepsilon(\boldsymbol{x}) \cap \operatorname{aff} C \subseteq C\}, \tag{2.5}$$

where $B_\varepsilon(\boldsymbol{x})$ is a ball with radius ε centered at $\boldsymbol{x}$ and $\operatorname{aff} C$ is the affine hull of C.

Definition 1 (Fenchel conjugate)**.** Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ be a proper, convex and lower semi-continuous function. Then, the function $f^* : \mathbb{R}^n \to \mathbb{R}_\infty$ defined by

$$f^*(\boldsymbol{\xi}) \coloneqq \sup_{\boldsymbol{x} \in \mathbb{R}^n} \langle \boldsymbol{x}, \boldsymbol{\xi} \rangle - f(\boldsymbol{x}) \tag{2.6}$$

is called the *Fenchel conjugate* of f.

Theorem 2 (Fenchel-Rockafellar duality)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ and $g : \mathbb{R}^m \to \mathbb{R}_\infty$ be proper, convex and lower semi-continuous functions and let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. It holds that*

$$\inf_{\boldsymbol{x} \in \mathbb{R}^n} \; f(\boldsymbol{x}) + g(\boldsymbol{A}\boldsymbol{x}) \;=\; \sup_{\boldsymbol{y} \in \mathbb{R}^m} \; -g^*(\boldsymbol{y}) - f^*(-\boldsymbol{A}^\top \boldsymbol{y}) \tag{2.7}$$

if either of the following conditions is satisfied:

1. *There exists an $\boldsymbol{x}_0 \in \operatorname{ri}(\operatorname{dom} f)$ such that $\boldsymbol{A}\boldsymbol{x}_0 \in \operatorname{ri}(\operatorname{dom} g)$.*
2. *There exists a $\boldsymbol{y}_0 \in \operatorname{ri}(\operatorname{dom} g^*)$ such that $\boldsymbol{A}^\top \boldsymbol{y}_0 \in \operatorname{ri}(\operatorname{dom} f^*)$.*

Under 1. the supremum is attained at some $\boldsymbol{y}$, while under 2. the infimum is attained at some $\boldsymbol{x}$.

Definition 3 (Subdifferential)**.** Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ be a convex function. A vector $\boldsymbol{\xi} \in \mathbb{R}^n$ is said to be a *subgradient* of f at a point $\boldsymbol{x} \in \mathbb{R}^n$ if and only if it holds that

$$\forall \boldsymbol{z} \in \mathbb{R}^n : f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{\xi}, \boldsymbol{z} - \boldsymbol{x} \rangle. \tag{2.8}$$

The set $\partial f(\boldsymbol{x})$ that contains all subgradients of f at $\boldsymbol{x}$ is called the *subdifferential* of f at $\boldsymbol{x}$.

Lemma 4. *Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ be a convex and differentiable function. Then, it holds for all $\boldsymbol{x} \in \mathbb{R}^n$ that*

$$\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}. \tag{2.9}$$

Proof. The condition

$$\forall \varepsilon > 0, \boldsymbol{z} \in \mathbb{R}^n : f(\boldsymbol{x} + \varepsilon \boldsymbol{z}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{\xi}, \varepsilon \boldsymbol{z} \rangle$$

is equivalent to (2.8). Since f is differentiable, it holds that

$$\lim_{\varepsilon \to 0} \frac{f(\boldsymbol{x} + \varepsilon \boldsymbol{z}) - f(\boldsymbol{x})}{\varepsilon} = \nabla_{\boldsymbol{z}} f(\boldsymbol{x}) = \langle \nabla f(\boldsymbol{x}), \boldsymbol{z} \rangle,$$

where $\nabla_{\boldsymbol{z}} f(\boldsymbol{x})$ is the directional derivative of f at $\boldsymbol{x}$ in direction $\boldsymbol{z}$. Consequently, each subgradient $\boldsymbol{\xi}$ satisfies

$$\forall \boldsymbol{z} \in \mathbb{R}^n : \langle \nabla f(\boldsymbol{x}) - \boldsymbol{\xi}, \boldsymbol{z} \rangle \geq 0$$

which is true if and only if $\boldsymbol{\xi} = \nabla f(\boldsymbol{x})$. □

Theorem 5 (Kuhn-Tucker conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ and $g : \mathbb{R}^m \to \mathbb{R}_\infty$ be proper, convex and lower semi-continuous functions and let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. In view of (2.7), it holds that the infimum and the supremum are attained at $\boldsymbol{x}^\star$ and $\boldsymbol{y}^\star$, respectively, if and only if the conditions*

$$-\boldsymbol{A}^\top \boldsymbol{y}^\star \in \partial f(\boldsymbol{x}^\star) \quad \text{and} \quad \boldsymbol{A}\boldsymbol{x}^\star \in \partial g^*(\boldsymbol{y}^\star) \tag{2.10}$$

are satisfied.

Now that we have two important results about duality in convex optimization at our disposal, our next step is to apply these results to the functions of our main interest. More precisely, we consider the problem of minimizing a function $f + g \circ \boldsymbol{A}$, where f is the ℓ_1-norm on $\mathbb{R}^n$, $\boldsymbol{A}$ is some linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ and g is the indicator function of the convex set $\{\boldsymbol{\zeta} \in \mathbb{R}^m : \|\boldsymbol{\zeta} - \boldsymbol{b}\|_\infty \leq \delta\}$ for some fixed $\boldsymbol{b} \in \mathbb{R}^m$ and $\delta \geq 0$.

Lemma 6. *Let $f : \mathbb{R}^n \to \mathbb{R}_\infty$ be defined by $f(\boldsymbol{x}) \coloneqq \|\boldsymbol{x}\|_1$. Then, it holds that*

$$f^*(\boldsymbol{\xi}) = I_{\|\cdot\|_\infty \leq 1}(\boldsymbol{\xi}) \quad \text{and} \quad \partial f(\boldsymbol{x}) = \{\boldsymbol{\xi} \in [-1, 1]^n \,|\, x_j \neq 0 \Rightarrow \xi_j = \operatorname{sign}(x_j)\}. \tag{2.11}$$

Proof. First of all, note that f is proper, convex and lower semi-continuous. According to Definition 1, we have

$$f^*(\boldsymbol{\xi}) = \sup_{\boldsymbol{x}\in\mathbb{R}^n} \langle \boldsymbol{x}, \boldsymbol{\xi}\rangle - \|\boldsymbol{x}\|_1.$$

If $|\xi_j| > 1$ for some j, then choosing $\operatorname{sign}(x_j) = \operatorname{sign}(\xi_j)$ and taking the limit $|x_j| \to \infty$ shows that $f^*(\boldsymbol{\xi}) = \infty$. Otherwise, it holds that $\langle \boldsymbol{x}, \boldsymbol{\xi}\rangle - \|\boldsymbol{x}\|_1 \leq 0$ and hence, we obtain $f^*(\boldsymbol{\xi}) = 0$ because the supremum is attained at $\boldsymbol{x} = \mathbf{0}$. It follows that the first statement is true. To see that the second one is true as well, note that (2.8) implies

$$\langle \boldsymbol{\xi}, \boldsymbol{x}\rangle - \|\boldsymbol{x}\|_1 \geq \sup_{\boldsymbol{z}\in\mathbb{R}^n} \langle \boldsymbol{\xi}, \boldsymbol{z}\rangle - \|\boldsymbol{z}\|_1 = f^*(\boldsymbol{\xi})$$

and thus, $\|\boldsymbol{\xi}\|_\infty \leq 1$ holds for each subgradient of f at $\boldsymbol{x}$. Now, suppose that $\xi_j \neq \operatorname{sign}(x_j)$ for any $x_j \neq 0$. As we have just shown that $\|\boldsymbol{\xi}\|_\infty \leq 1$, we conclude that $\|\boldsymbol{x}\|_1 - \langle \boldsymbol{\xi}, \boldsymbol{x}\rangle > 0$ which shows that (2.8) is not satisfied for $\boldsymbol{z} = \mathbf{0}$. Hence, we have $\xi_j = \operatorname{sign}(x_j)$ and $\|\boldsymbol{x}\|_1 - \langle \boldsymbol{\xi}, \boldsymbol{x}\rangle = 0$. It follows that the second statement is true, since $\|\boldsymbol{z}\|_1 \geq \langle \boldsymbol{\xi}, \boldsymbol{z}\rangle$ holds whenever $\|\boldsymbol{\xi}\|_\infty \leq 1$. □

Lemma 7. *Let $h : \mathbb{R}^m \to \mathbb{R}_\infty$ be a proper, convex and lower semi-continuous function, $\boldsymbol{b} \in \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}_\infty$ be defined by $g(\boldsymbol{\zeta}) \coloneqq h(\boldsymbol{\zeta} - \boldsymbol{b})$. Then, the Fenchel conjugate of g can be written as*

$$g^*(\boldsymbol{y}) = h^*(\boldsymbol{y}) + \langle \boldsymbol{b}, \boldsymbol{y}\rangle. \tag{2.12}$$

Proof. Using Definition 1 and substituting $\boldsymbol{\zeta}' \coloneqq \boldsymbol{\zeta} - \boldsymbol{b}$, we obtain

$$\begin{aligned} g^*(\boldsymbol{y}) &= \sup_{\boldsymbol{\zeta}\in\mathbb{R}^m} \langle \boldsymbol{\zeta}, \boldsymbol{y}\rangle - h(\boldsymbol{\zeta} - \boldsymbol{b}) \\ &= \sup_{\boldsymbol{\zeta}'\in\mathbb{R}^m} \langle \boldsymbol{\zeta}' + \boldsymbol{b}, \boldsymbol{y}\rangle - h(\boldsymbol{\zeta}') \\ &= \Big(\sup_{\boldsymbol{\zeta}'\in\mathbb{R}^m} \langle \boldsymbol{\zeta}', \boldsymbol{y}\rangle - h(\boldsymbol{\zeta}')\Big) + \langle \boldsymbol{b}, \boldsymbol{y}\rangle \\ &= h^*(\boldsymbol{y}) + \langle \boldsymbol{b}, \boldsymbol{y}\rangle. \end{aligned}$$

□

Lemma 8. *Let $\delta \geq 0$. Then, it holds that*

$$I^*_{\|\cdot\|_\infty \leq \delta}(\boldsymbol{y}) = \delta\|\boldsymbol{y}\|_1. \tag{2.13}$$

Proof. First, note that each indicator function I_C associated with a non-empty and convex set C is proper, convex and lower semi-continuous. We use Definition 1 again and get

$$\begin{aligned} I^*_{\|\cdot\|_\infty \leq \delta}(\boldsymbol{y}) &= \sup_{\boldsymbol{\zeta}\in\mathbb{R}^m} \langle \boldsymbol{\zeta}, \boldsymbol{y}\rangle - I_{\|\cdot\|_\infty \leq \delta}(\boldsymbol{\zeta}) \\ &= \sup_{\boldsymbol{\zeta}\in\mathbb{R}^m} \langle \boldsymbol{\zeta}, \boldsymbol{y}\rangle \quad \text{s.t. } \|\boldsymbol{\zeta}\|_\infty \leq \delta \\ &= \delta\|\boldsymbol{y}\|_1. \end{aligned}$$

□

Corollary 9. *Let $\delta \geq 0$ and $\boldsymbol{b} \in \mathbb{R}^m$. Then, it holds that*

$$I^*_{\|\cdot-\boldsymbol{b}\|_\infty \leq \delta}(\boldsymbol{y}) = \delta\|\boldsymbol{y}\|_1 + \langle \boldsymbol{b}, \boldsymbol{y}\rangle. \tag{2.14}$$

Proof. The statement follows as a direct consequence of Lemmas 7 and 8. □

Corollary 10. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\delta \geq 0$. Then, it holds that*

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 + I_{\|\cdot-b\|_\infty\leq\delta}(\boldsymbol{A}\boldsymbol{x}) \\ =\ &\sup_{\boldsymbol{y}\in\mathbb{R}^m} \ -\delta\|\boldsymbol{y}\|_1 - \langle \boldsymbol{b}, \boldsymbol{y}\rangle - I_{\|\cdot\|_\infty\leq 1}(\boldsymbol{A}^\top \boldsymbol{y})\,. \end{aligned} \tag{2.15}$$

Proof. With $f = \|\cdot\|_1$ and $g = I_{\|\cdot-b\|_\infty\leq\delta}$, both f and g are proper, convex and lower semi-continuous. Lemma 6 states that $f^* = I_{\|\cdot\|_\infty\leq 1}$ and by Corollary 9, it holds that $g^* = \delta\|\cdot\|_1 + \langle \boldsymbol{b}, \cdot\rangle$. As $\operatorname{ri}(\operatorname{dom} g^*) = \mathbb{R}^n$ and $\operatorname{ri}(\operatorname{dom} f^*) = \{\|\boldsymbol{\xi}\|_\infty < 1\}$, the second condition in Theorem 2 is satisfied at $\boldsymbol{y}_0 = \boldsymbol{0}$. Hence, there exists an $\boldsymbol{x} \in \mathbb{R}^n$ where the infimum is attained. □

Corollary 11. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\delta \geq 0$. Then, the minimum* and *the supremum in* (2.15) *are attained at $\boldsymbol{x}^\star$ and $\boldsymbol{y}^\star$, respectively, if and only if*

$$-\boldsymbol{A}^\top \boldsymbol{y}^\star \in \partial\|\boldsymbol{x}^\star\|_1 \quad \text{and} \quad \boldsymbol{A}\boldsymbol{x}^\star - \boldsymbol{b} \in \delta\partial\|\boldsymbol{y}^\star\|_1. \tag{2.16}$$

Proof. We apply Theorem 5 with f and g as in the proof of Corollary 10. Additionally, we use [38, Theorem 23.8] and Lemma 4 to see that $\partial(\delta\|\boldsymbol{y}\|_1 + \langle \boldsymbol{b}, \boldsymbol{y}\rangle) = \delta\partial\|\boldsymbol{y}\|_1 + \boldsymbol{b}$. □

Theorem 12. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\delta \geq 0$ such that there exists an $\boldsymbol{x}_0 \in \mathbb{R}^n$ satisfying $\|\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}\|_\infty \leq \delta$. Then, it holds that*

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \delta \\ =\ &\max_{\boldsymbol{y}\in\mathbb{R}^m} \ -\delta\|\boldsymbol{y}\|_1 - \langle \boldsymbol{b}, \boldsymbol{y}\rangle \quad \text{s.t } \|\boldsymbol{A}^\top \boldsymbol{y}\|_\infty \leq 1 \end{aligned} \tag{2.17}$$

and the minimum and the maximum are attained at $\boldsymbol{x}^\star$ and $\boldsymbol{y}^\star$ if and only if

$$-\boldsymbol{A}^\top \boldsymbol{y}^\star \in \partial\|\boldsymbol{x}^\star\|_1 \quad \text{and} \quad \boldsymbol{A}\boldsymbol{x}^\star - \boldsymbol{b} \in \delta\partial\|\boldsymbol{y}^\star\|_1. \tag{2.18}$$

Proof. We exploit that both problems in (2.17) have respective equivalent formulations as linear programs (as to that, we anticipate Lemma 45) which are dual to each other in the sense of LP duality. The minimization problem is feasible by assumption. It is further bounded due to $\|\cdot\|_1 \geq 0$. Therefore, the *Duality Theorem of Linear Programming* (see, e.g., [29]) states that the maximization problem is feasible and bounded as well and that both problems attain the same optimal objective function value. Therewith, the final part of the statement follows from Corollary 11. □

3 Homotopy Method

In this chapter, we introduce our homotpy method. To that end, we first revisit the primal-dual optimality conditions for (P_δ) in Section 3.1, and show that they can be formulated in terms of linear equations and inequalities if either the primal or the dual variable are fixed. Afterwards, in Section 3.2, we use this property in order to derive an alternating primal-dual update scheme. In Section 3.3, we establish a theorem of the alternative which is the key result for our analysis in Section 3.4, where we show that the proposed method terminates after a finite number of iterations yielding an optimal solution of (P_δ). Up to this point, all mentioned results have appeared in [5], with the involvement of the author. However, the analysis of the solution path of (P_δ) in Section 3.5 and the upper and lower complexity bounds given in Sections 3.6 and 3.7 are novel and first appearing in this thesis.

3.1 Optimality Conditions

We consider the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_\infty \leq \delta, \tag{P_δ}$$

with $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{b}\in\mathbb{R}^m$ and $\delta\geq 0$ assuming that there exists an $\boldsymbol{x}_0\in\mathbb{R}^n$ such that $\|\boldsymbol{A}\boldsymbol{x}_0-\boldsymbol{b}\|_\infty\leq\delta$. Theorem 12 states that $\boldsymbol{x}^\star$ is an optimal solution of (P_δ) if and only if there exists a $\boldsymbol{y}^\star\in\mathbb{R}^m$ such that the conditions

$$-\boldsymbol{A}^\top\boldsymbol{y}^\star\in\partial\|\boldsymbol{x}^\star\|_1 \quad \text{and} \quad \boldsymbol{A}\boldsymbol{x}^\star-\boldsymbol{b}\in\delta\partial\|\boldsymbol{y}^\star\|_1 \tag{3.1}$$

are satisfied. Each such $\boldsymbol{y}^\star$ is by construction an optimal solution to the dual problem of (P_δ), which is

$$\max_{\boldsymbol{y}\in\mathbb{R}^m} \ -\boldsymbol{b}^\top\boldsymbol{y}-\delta\|\boldsymbol{y}\|_1 \quad \text{s.t. } \|\boldsymbol{A}^\top\boldsymbol{y}\|_\infty\leq 1. \tag{D_δ}$$

Therefore, we sometimes refer to $\boldsymbol{x}^\star$ as a *primal solution*, to $\boldsymbol{y}^\star$ as a *dual solution* and to $(\boldsymbol{x}^\star,\boldsymbol{y}^\star)$ as an *optimal pair*. For a thorough understanding of the conditions (3.1), it is helpful to define the sets

$$\begin{aligned} &S\coloneqq\{j : x_j^\star\neq 0\}, && W\coloneqq\{i : |\boldsymbol{a}_i^\top\boldsymbol{x}^\star-b_i|=\delta\},\\ &(\textit{primal support}) && (\textit{primal active set})\\[2ex] &\Sigma\coloneqq\{j : |\boldsymbol{A}_j^\top\boldsymbol{y}^\star|=1\}, && \Omega\coloneqq\{i : y_i^\star\neq 0\}.\\ &(\textit{dual active set}) && (\textit{dual support}) \end{aligned}$$

Since $\partial\|\boldsymbol{x}^\star\|_1 = \{\boldsymbol{\xi} \in [-1,1]^n : \boldsymbol{\xi}_S = \operatorname{sign}(\boldsymbol{x}^\star_S)\}$, the conditions (3.1) require that $S \subseteq \Sigma$ and $\Omega \subseteq W$. Moreover, the partitioned conditions

$$\begin{aligned} -\boldsymbol{A}_S^\top \boldsymbol{y}^\star &= \operatorname{sign}(\boldsymbol{x}^\star_S) & \boldsymbol{A}^\Omega \boldsymbol{x}^\star - \boldsymbol{b}_\Omega &= \delta \operatorname{sign}(\boldsymbol{y}^\star_\Omega) \\ -\mathbb{1} \leq -\boldsymbol{A}_{S^c}^\top \boldsymbol{y}^\star &\leq \mathbb{1} & -\delta\mathbb{1} \leq \boldsymbol{A}^{\Omega^c} \boldsymbol{x}^\star - \boldsymbol{b}_{\Omega^c} &\leq \delta\mathbb{1} \end{aligned} \tag{C_δ}$$

are equivalent to (3.1).

3.2 Homotopy Approach

Our approach is to solve a sequence of problems $(\mathrm{P}_{\delta^k})_{k=0,\dots,K}$, where

$$\delta_0 > \delta_1 > \cdots > \delta_K = \delta.$$

We assume that an optimal pair $(\boldsymbol{x}^0, \boldsymbol{y}^0)$ for (P_{δ^0}) is known a priori, while the number K of subsequent problems as well as the associated optimal pairs are initially unknown. The underlying motivation is that the transition from an optimal pair $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ for (P_{δ^k}) to a subsequent optimal pair $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ for $(\mathrm{P}_{\delta^{k+1}})$ can be much less complex than solving (P_δ) directly.

The main idea behind the iterations of our method is the following: Suppose that $\delta^k > \delta$ and that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal pair for (P_{δ^k}). In order to find δ^{k+1} and an associated optimal pair $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$, we proceed in two steps. First, we identify a new dual solution $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$ such that $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ is still an optimal pair for (P_{δ^k}). In a second step, we construct a new primal solution $\boldsymbol{x}^{k+1} \neq \boldsymbol{x}^k$ and a $t^{k+1} > 0$ such that $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ is an optimal pair for $(\mathrm{P}_{\delta^k - t^{k+1}})$, before we finally set $\delta^{k+1} \coloneqq \delta^k - t^{k+1}$.

In the following, we propose computing $\boldsymbol{x}^{k+1}$, $\boldsymbol{y}^{k+1}$ and δ^{k+1} by solving two relatively small linear programs, where the constraints are derived from the conditions (C_{δ^k}) and $(\mathrm{C}_{\delta^{k+1}})$, respectively. Afterwards, we prove a theorem of the alternative which is key to showing that the proposed method terminates after a finite number of iterations, yielding an optimal pair $(\boldsymbol{x}^\star, \boldsymbol{y}^\star)$ for (P_δ).

Analogous to above, we define the sets

$$\begin{aligned} S_k &\coloneqq \{j : x_j^k \neq 0\}, & W_k &\coloneqq \{i : |\boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i| = \delta^k\}, \\ \Sigma_k &\coloneqq \{j : |\boldsymbol{A}_j^\top \boldsymbol{y}^k| = 1\}, & \Omega_k &\coloneqq \{i : y_i^k \neq 0\}, \end{aligned}$$

relating to the iterates $\boldsymbol{x}^k$ and $\boldsymbol{y}^k$.

3.2.1 Dual Updates

Suppose that $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal pair for (P_{δ^k}) and hence, the conditions (C_{δ^k}) are valid for $\boldsymbol{x}^k$ and $\boldsymbol{y}^k$. In order to determine a new dual solution, we fix $\boldsymbol{x}^k$ in (C_{δ^k}) and search a $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$ such that the conditions stay valid. While (C_{δ^k}) still involves non-linear conditions, the following lemma provides an equivalent linear reformulation.

Lemma 13. *Let $\boldsymbol{x}^k$ be an optimal solution of (P_{δ^k}). Then, $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ is an optimal pair for (P_{δ^k}) if and only if $\boldsymbol{y}^{k+1}$ is a solution of*

$$\begin{aligned} -\boldsymbol{A}_{S_k}^\top \boldsymbol{y} &= \operatorname{sign}(\boldsymbol{x}_{S_k}^k) \\ -\mathbb{1} \leq -\boldsymbol{A}_{S_k^c}^\top \boldsymbol{y} &\leq \mathbb{1} \\ -\operatorname{sign}(\boldsymbol{A}^{W_k}\boldsymbol{x}^k - \boldsymbol{b}_{W_k}) \odot \boldsymbol{y}_{W_k} &\leq \boldsymbol{0} \\ \boldsymbol{y}_{W_k^c} &= \boldsymbol{0}. \end{aligned} \tag{C_D^k}$$

Proof. The statement follows from a direct comparison of (C_D^k) and (C_{δ^k}) with $\boldsymbol{x}^k$ fixed. While the first two conditions in (C_D^k) correspond exactly with the respective conditions in (C_{δ^k}), the remaining two conditions in (C_D^k) are equivalent to $\boldsymbol{A}^\Omega \boldsymbol{x}^k - \boldsymbol{b}_\Omega = \delta^k \operatorname{sign}(\boldsymbol{y}_\Omega)$. The latter equation implies that $\Omega \subseteq W_k$. Hence, it follows that $\boldsymbol{y}_{W_k^c} = \boldsymbol{0}$ and, in combination with the same equation, that $-\operatorname{sign}(\boldsymbol{A}^{W_k}\boldsymbol{x}^k - \boldsymbol{b}_{W_k}) \odot \boldsymbol{y}_{W_k} \leq \boldsymbol{0}$. Vice versa, it follows from $\boldsymbol{y}_{W_k^c} = \boldsymbol{0}$ that $\Omega \subseteq W_k$. Therewith, we obtain $|\boldsymbol{A}^\Omega \boldsymbol{x}^k - \boldsymbol{b}_\Omega| = \delta^k \mathbb{1}$ and the third condition in (C_D^k) ensures that $\boldsymbol{A}^\Omega \boldsymbol{x}^k - \boldsymbol{b}_\Omega = \delta^k \operatorname{sign}(\boldsymbol{y}_\Omega)$. The remaining condition in (C_{δ^k}) does not depend on $\boldsymbol{y}$ because $\boldsymbol{x}^k$ is in particular feasible for (P_{δ^k}). □

As the previous iterate $\boldsymbol{y}^k$ is always feasible for (C_D^k), this holds as well for each convex combination of $\boldsymbol{y}^k$ and $\boldsymbol{y}^{k+1}$. Thus, provided that one solution $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$ exists, (C_D^k) has infinitely many solutions. To overcome this issue, we propose to choose a suitable linear objective function $\boldsymbol{\psi} \in \mathbb{R}^m$ and perform the update

$$\boldsymbol{y}^{k+1} \in \operatorname*{arg\,min}_{\boldsymbol{y} \in \mathbb{R}^m} \boldsymbol{\psi}^\top \boldsymbol{y} \quad \text{s.t. } \boldsymbol{y} \text{ satisfies } (\mathrm{C}_D^k) \tag{U_D^k}$$

which corresponds to solving a linear program with $|W_k|$ bounded variables and $2n - |S_k|$ constraints. We postpone the question how to choose $\boldsymbol{\psi}$ to the next section (cf. Theorem 20).

3.2.2 Primal Updates

After the dual update, we have an optimal pair $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ for (P_{δ^k}) at hand. We introduce an auxiliary non-negative variable t representing the local decrease of the homotopy parameter. Then, we fix $\boldsymbol{y}^{k+1}$ in ($\mathrm{C}_{\delta^k - t}$) and seek for a $\boldsymbol{x}^{k+1} \neq \boldsymbol{x}^k$ and a $t^{k+1} > 0$ such that the conditions stay satisfied.

Lemma 14. *Let $\boldsymbol{y}^{k+1}$ be an optimal solution of (D_{δ^k}). Then, $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ is an optimal pair for ($P_{\delta^k - t^{k+1}}$) with $\delta^k - t^{k+1} \geq \delta$ if and only if $\boldsymbol{x}^{k+1}$ and t^{k+1} form a solution of*

$$\begin{aligned} \boldsymbol{A}^{\Omega_{k+1}}\boldsymbol{x} - \boldsymbol{b}_{\Omega_{k+1}} &= (\delta^k - t)\operatorname{sign}(\boldsymbol{y}_{\Omega_{k+1}}^{k+1}) \\ -(\delta^k - t)\mathbb{1} \leq \boldsymbol{A}^{\Omega_{k+1}^c}\boldsymbol{x} - \boldsymbol{b}_{\Omega_{k+1}^c} &\leq (\delta^k - t)\mathbb{1} \\ \boldsymbol{A}_{\Sigma_{k+1}}^\top \boldsymbol{y}^{k+1} \odot \boldsymbol{x}_{\Sigma_{k+1}} &\leq \boldsymbol{0} \\ \boldsymbol{x}_{\Sigma_{k+1}^c} &= \boldsymbol{0} \\ t &\leq \delta^k - \delta. \end{aligned} \tag{C_P^k}$$

Proof. Analogous to above, the statement follows from a direct comparison of (C_P^k) and $(\mathrm{C}_{\delta^k - t})$, this time with $\boldsymbol{y}^{k+1}$ fixed. The first two conditions in (C_P^k) comply with the respective conditions in $(\mathrm{C}_{\delta^k - t})$ and the condition $-\mathbb{1} \leq -\boldsymbol{A}_{S^c}^\top \boldsymbol{y}^{k+1} \leq \mathbb{1}$ in $(\mathrm{C}_{\delta^k - t})$ does not depend on $\boldsymbol{x}$ because $\boldsymbol{y}^{k+1}$ is feasible for (D_{δ^k}). Completely analogous to the argumentation in the proof of Lemma 13, it follows further that the third and fourth conditions in (C_P^k) are equivalent to the condition $-\boldsymbol{A}_S^\top \boldsymbol{y}^{k+1} = \operatorname{sign}(\boldsymbol{x}_S)$ in $(\mathrm{C}_{\delta^k - t})$. □

The last condition in (C_P^k) solely prevents us from jumping over an optimal solution of the original problem (P_δ). Analogous to the situation in the dual update, (C_P^k) has infinitely many solutions, provided that at least one solution distinct from $(\boldsymbol{x}^k, 0)$ exists. Therefore, we propose to perform the update

$$(\boldsymbol{x}^{k+1}, t^{k+1}) \in \underset{(\boldsymbol{x},t)\in\mathbb{R}^n\times\mathbb{R}}{\arg\max} \; t \quad \text{s.t. } (\boldsymbol{x}, t) \text{ satisfies } (\mathrm{C}_P^k) \tag{U_P^k}$$

and $\delta^{k+1} \coloneqq \delta^k - t^{k+1}$. This update can be considered intuitive as it maximizes the local decrease of the homotopy parameter. It amounts to solving a linear program with $|\Sigma_{k+1}| + 1$ bounded variables and $2m - |\Omega_{k+1}|$ constraints.

3.3 A Theorem of the Alternative

In the previous section, we derived two sets of linear equations and inequalities that serve as constraints in the linear programs to determine the next dual and primal iterate, respectively. When formulating the dual update, we left open the question how to choose the linear objective function $\boldsymbol{\psi}$. Naturally, we look for a function such that alternating updates in the form (U_D^k) and (U_P^k) result in a convergent method, yielding an optimal pair for our original problem. We proceed in two steps in order to establish our final choice $\boldsymbol{\psi} = -\operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b})$. First, we show that both the dual and the primal update can be expressed in terms of *improvement directions*. Second, we prove that a dual improvement direction with respect to $\boldsymbol{\psi}$ exists if and only if there exists no primal improvement direction. As a consequence, we obtain a theorem of the alternative that encompasses both the existence of improvement directions and the progress that can be made by performing dual and primal updates, respectively.

For the moment, we change our point of view from the iterates of our method to an arbitrary optimal pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ for $(\mathrm{P}_{\hat{\delta}})$ for some $\hat{\delta} > \delta$. Throughout, we refer to the corresonding supports and active sets as S, W, Σ and Ω. This more general approach will become useful subsequently, when we examine the set of solutions for all possible values of the homotopy parameter.

Lemma 15. *Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an optimal pair for $(P_{\hat{\delta}})$ for some $\hat{\delta} > \delta$, and let $\boldsymbol{\psi} \in \mathbb{R}^m$.*

Then, the system

$$\begin{aligned} \boldsymbol{\psi}^\top \boldsymbol{e} &< 0 && \text{(3.2a)}\\ \boldsymbol{A}_S^\top \boldsymbol{e} &= \mathbf{0} && \text{(3.2b)}\\ \boldsymbol{A}_{\Sigma\setminus S}^\top \hat{\boldsymbol{y}} \odot \boldsymbol{A}_{\Sigma\setminus S}^\top \boldsymbol{e} &\leq \mathbf{0} && \text{(3.2c)}\\ -\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{e}_{W\setminus\Omega} &\leq \mathbf{0} && \text{(3.2d)}\\ \boldsymbol{e}_{W^c} &= \mathbf{0} && \text{(3.2e)} \end{aligned}$$

is feasible if and only if $\hat{\boldsymbol{y}}$ *is not an optimal solution of*

$$\begin{aligned} \min_{\boldsymbol{y}\in\mathbb{R}^m} \quad & \boldsymbol{\psi}^\top \boldsymbol{y} && && \text{(3.3a)}\\ \text{s.t.} \quad & -\boldsymbol{A}_S^\top \boldsymbol{y} &&= \operatorname{sign}(\hat{\boldsymbol{x}}_S) && \text{(3.3b)}\\ & -\mathbb{1} \leq -\boldsymbol{A}_{S^c}^\top \boldsymbol{y} &&\leq \mathbb{1} && \text{(3.3c)}\\ & -\operatorname{sign}(\boldsymbol{A}^W\hat{\boldsymbol{x}} - \boldsymbol{b}_W) \odot \boldsymbol{y}_W &&\leq \mathbf{0} && \text{(3.3d)}\\ & \boldsymbol{y}_{W^c} &&= \mathbf{0}. && \text{(3.3e)} \end{aligned}$$

If $\hat{\boldsymbol{e}}$ *is a solution of (3.2), then there exists an* $\hat{s} > 0$ *such that* $\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}$ *is feasible for* (3.3) *and* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}})$ *is an optimal pair for* $(P_{\hat{\delta}})$ *for all* $0 \leq s \leq \hat{s}$*. Conversely, if* $(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ *is an optimal pair for* $(P_{\hat{\delta}})$ *and* $\boldsymbol{\psi}^\top\tilde{\boldsymbol{y}} < \boldsymbol{\psi}^\top\hat{\boldsymbol{y}}$*, then* $\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}}$ *is feasible for (3.2) and* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} + s\tilde{\boldsymbol{e}})$ *is an optimal pair for* $(P_{\hat{\delta}})$ *for all* $0 \leq s \leq 1$*.*

Proof. Since $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\hat{\delta}})$, we conclude from Lemma 13 that $\hat{\boldsymbol{y}}$ is feasible for (3.3). In particular, it holds that

$$\begin{aligned} -\boldsymbol{A}_S^\top\hat{\boldsymbol{y}} &= \operatorname{sign}(\hat{\boldsymbol{x}}_S),\\ |-\boldsymbol{A}_{\Sigma\setminus S}^\top\hat{\boldsymbol{y}}| &= \mathbb{1},\\ -\mathbb{1} < -\boldsymbol{A}_{\Sigma^c}^\top\hat{\boldsymbol{y}} &< \mathbb{1},\\ -\operatorname{sign}(\boldsymbol{A}^\Omega\hat{\boldsymbol{x}} - \boldsymbol{b}_\Omega) \odot \hat{\boldsymbol{y}}_\Omega &< \mathbf{0},\\ \hat{\boldsymbol{y}}_{W\setminus\Omega} &= \mathbf{0}. \end{aligned}$$

First, suppose that the system (3.2) is feasible and that $\hat{\boldsymbol{e}}$ is a corresponding solution. For arbitrary $s > 0$, it follows from (3.2b) and (3.2e) that $\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}$ satisfies (3.3b) and (3.3e), respectively. Furthermore, (3.2c) ensures that there exists an $s_1 > 0$ such that $\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}$ still satisfies (3.3c) for $0 \leq s \leq s_1$, and (3.2d) ensures that there exists an $s_2 > 0$ such that $\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}$ obeys (3.3d) for $0 \leq s \leq s_2$. Thus, we can choose $\hat{s} \coloneqq \min\{s_1, s_2\}$ and obtain that $\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}$ is feasible for (3.3b)–(3.3e) for all $0 \leq s \leq \hat{s}$. By Lemma 13, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}})$ is an optimal pair for the respective values of s. Moreover, because of (3.2a), it holds that $\boldsymbol{\psi}^\top(\hat{\boldsymbol{y}} + s\hat{\boldsymbol{e}}) < \boldsymbol{\psi}^\top\hat{\boldsymbol{y}}$, which shows that $\hat{\boldsymbol{y}}$ is *not* a minimizer in (3.3).

Now, suppose that $(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is an optimal pair with $\boldsymbol{\psi}^\top\tilde{\boldsymbol{y}} < \boldsymbol{\psi}^\top\hat{\boldsymbol{y}}$ and consider $\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}}$. The conditions (3.3b)–(3.3e) continue to hold for $\hat{\boldsymbol{y}} + \tilde{\boldsymbol{e}}$, which implies that $\tilde{\boldsymbol{e}}$ satisfies (3.2b)–(3.2e). Additionally, it holds that $\boldsymbol{\psi}^\top(\hat{\boldsymbol{y}} + \tilde{\boldsymbol{e}}) < \boldsymbol{\psi}^\top\hat{\boldsymbol{y}}$, which shows that $\tilde{\boldsymbol{e}}$ obeys

(3.2a). Thus, $\tilde{\boldsymbol{e}}$ is feasible for (3.2). As both $\tilde{\boldsymbol{y}}$ and $\hat{\boldsymbol{y}}$ satisfy (3.3b)–(3.3e), each convex combination $\hat{\boldsymbol{y}} + s(\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}})$ is feasible as well, which shows that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} + s\tilde{\boldsymbol{e}})$ is an optimal pair for all $0 \leq s \leq 1$. □

Example 16. We consider the case $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^k, \boldsymbol{y}^k)$ in which the dual update (U_D^k) corresponds exactly to (3.3). Lemma 15 states that there exists an improvement direction according to (3.2) if and only if $\boldsymbol{y}^k$ is not an optimal solution of (3.3). Consequently, in that case, it holds that $\boldsymbol{\psi}^\top \boldsymbol{y}^{k+1} < \boldsymbol{\psi}^\top \boldsymbol{y}^k$ and thus, $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$. The particular improvement direction corresponding to the dual update is $\boldsymbol{e}^{k+1} = \boldsymbol{y}^{k+1} - \boldsymbol{y}^k$. For each $0 \leq s \leq 1$, the convex combination $\boldsymbol{y}^k + s\boldsymbol{e}^{k+1}$ induces an optimal pair for (P_{δ^k}) as well. Note that we would not obtain further optimal pairs for $s > 1$, because this would contradict the optimality of $\boldsymbol{y}^{k+1}$ in (3.3).

Lemma 17. *Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an optimal pair for $(P_{\hat{\delta}})$ for some $\hat{\delta} > \delta$. Then, the system*

$$\begin{aligned}
\boldsymbol{A}^{\Omega}\boldsymbol{d} &= -\operatorname{sign}(\hat{\boldsymbol{y}}_\Omega) && \text{(3.4a)}\\
\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}^{W\setminus\Omega}\boldsymbol{d} &\leq -\mathbb{1} && \text{(3.4b)}\\
\boldsymbol{A}_{\Sigma\setminus S}^\top \hat{\boldsymbol{y}} \odot \boldsymbol{d}_{\Sigma\setminus S} &\leq \boldsymbol{0} && \text{(3.4c)}\\
\boldsymbol{d}_{\Sigma^c} &= \boldsymbol{0} && \text{(3.4d)}
\end{aligned}$$

is feasible if and only if $(\hat{\boldsymbol{x}}, 0)$ is not an optimal solution of

$$\begin{aligned}
\max_{(\boldsymbol{x},t)\in\mathbb{R}^n\times\mathbb{R}} \quad & t && && \text{(3.5a)}\\
\text{s.t} \quad & \boldsymbol{A}^{\Omega}\boldsymbol{x} - \boldsymbol{b}_\Omega = (\hat{\delta} - t)\operatorname{sign}(\hat{\boldsymbol{y}}_\Omega) && && \text{(3.5b)}\\
-(\hat{\delta} - t)\mathbb{1} \leq \boldsymbol{A}^{\Omega^c}\boldsymbol{x} - \boldsymbol{b}_{\Omega^c} &\leq (\hat{\delta} - t)\mathbb{1} && && \text{(3.5c)}\\
\boldsymbol{A}_\Sigma^\top \hat{\boldsymbol{y}} \odot \boldsymbol{x}_\Sigma &\leq \boldsymbol{0} && && \text{(3.5d)}\\
\boldsymbol{x}_{\Sigma^c} &= \boldsymbol{0} && && \text{(3.5e)}\\
t &\leq \hat{\delta} - \delta. && && \text{(3.5f)}
\end{aligned}$$

If $\hat{\boldsymbol{d}}$ is a solution of (3.4), then there exists a $\hat{t} > 0$ such that $\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}$ is feasible for (3.5) and $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(P_{\hat{\delta}-t})$ for all $0 \leq t \leq \hat{t}$. Conversely, if $(\tilde{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(P_{\tilde{\delta}})$ and $\tilde{\delta} < \hat{\delta}$, then $\tilde{\boldsymbol{d}} = (\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}})/(\hat{\delta} - \tilde{\delta})$ is feasible for (3.4) and $(\hat{\boldsymbol{x}} + t\tilde{\boldsymbol{d}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(P_{\hat{\delta}-t})$ for all $0 \leq t \leq \hat{\delta} - \tilde{\delta}$.

Proof. Since $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\hat{\delta}})$, it holds that $(\hat{\boldsymbol{x}}, 0)$ is feasible for (3.5). In particular, we obtain

$$\begin{aligned}
\boldsymbol{A}^{\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_\Omega &= \hat{\delta}\operatorname{sign}(\hat{\boldsymbol{y}}_\Omega),\\
|\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}| &= \hat{\delta}\mathbb{1},\\
-\hat{\delta}\mathbb{1} < \boldsymbol{A}^{W^c}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W^c} &< \hat{\delta}\mathbb{1},\\
\boldsymbol{A}_S^\top \hat{\boldsymbol{y}} \odot \hat{\boldsymbol{x}}_S &< \boldsymbol{0},\\
\hat{\boldsymbol{x}}_{\Sigma\setminus S} &= \boldsymbol{0}.
\end{aligned}$$

Suppose that the system (3.4) is feasible and that $\hat{\boldsymbol{d}}$ is a corresponding solution. As $\hat{\boldsymbol{d}}$ fulfills (3.4a) and (3.4d), we get that for each $t > 0$, $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, t)$ fulfills (3.5b) and (3.5e). From (3.4b), we obtain that there exists a $t_1 > 0$ such that $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, t)$ satisfies (3.5c) for all $0 \leq t \leq t_1$. Because of (3.4c), there exists a $t_2 > 0$ such that $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, t)$ fulfills (3.5d) for all $0 \leq t \leq t_2$. Thus, we can choose $\hat{t} \coloneqq \min\{t_1, t_2, \hat{\delta} - \delta\} > 0$ and obtain that $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, t)$ is feasible for (3.5) for $0 \leq t \leq \hat{t}$. This shows that $(\hat{\boldsymbol{x}}, 0)$ is not an optimal solution of (3.5) and further, due to Lemma 14, that $(\hat{\boldsymbol{x}} + t\hat{\boldsymbol{d}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\hat{\delta}-t})$ for $0 \leq t \leq \hat{t}$.

Conversely, suppose that $(\hat{\boldsymbol{x}}, 0)$ is not an optimal solution of (3.5). Then, there exists a pair $(\tilde{\boldsymbol{x}}, \tilde{t})$ with $\tilde{t} > 0$ that satisfies (3.5b)–(3.5f), or equivalently, $(\tilde{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\tilde{\delta}})$ with $\tilde{\delta} = \hat{\delta} - \tilde{t}$. We see that $\tilde{\boldsymbol{d}} = (\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}})/\tilde{t}$ obeys (3.4a) and (3.4d). Further, it holds that

$$
\begin{aligned}
& \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}^{W\setminus\Omega}\tilde{\boldsymbol{d}} \\
= \;& \tfrac{1}{\tilde{t}} \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot ([\boldsymbol{A}^{W\setminus\Omega}\tilde{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}] - [\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}]) \\
= \;& \tfrac{1}{\tilde{t}} \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\tilde{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot |\boldsymbol{A}^{W\setminus\Omega}\tilde{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}| \\
& -\tfrac{1}{\tilde{t}} \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot |\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}| \\
\leq \;& \tfrac{1}{\tilde{t}} |\boldsymbol{A}^{W\setminus\Omega}\tilde{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}| - \tfrac{1}{\tilde{t}} |\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}| \;\leq\; \tfrac{1}{\tilde{t}} (\hat{\delta} - \tilde{t})\mathbb{1} - \tfrac{1}{\tilde{t}} \hat{\delta}\mathbb{1} \;=\; -\mathbb{1},
\end{aligned}
$$

so $\tilde{\boldsymbol{d}}$ satisfies (3.4b) as well. Eventually, (3.4c) also holds true, since

$$
\boldsymbol{A}^{\top}_{\Sigma\setminus S}\hat{\boldsymbol{y}} \odot \tilde{\boldsymbol{d}}_{\Sigma\setminus S} = \tfrac{1}{\tilde{t}} \underbrace{\boldsymbol{A}^{\top}_{\Sigma\setminus S}\hat{\boldsymbol{y}} \odot \tilde{\boldsymbol{x}}_{\Sigma\setminus S}}_{\leq \mathbf{0}} - \tfrac{1}{\tilde{t}} \boldsymbol{A}^{\top}_{\Sigma\setminus S}\hat{\boldsymbol{y}} \odot \underbrace{\hat{\boldsymbol{x}}_{\Sigma\setminus S}}_{=\mathbf{0}} \leq \mathbf{0}.
$$

We conclude that $\tilde{d}$ is feasible for (3.4). Moreover, since both $(\tilde{\boldsymbol{x}}, \tilde{t})$ and $(\hat{\boldsymbol{x}}, 0)$ satisfy (3.5b)–(3.5f), each convex combination $(\hat{\boldsymbol{x}} + (t/\tilde{t})(\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}), t)$ with $0 \leq t \leq \tilde{t}$ is feasible as well. Hence, $(\hat{\boldsymbol{x}} + t\tilde{\boldsymbol{d}}, \hat{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\hat{\delta}-t})$ for all those t. □

Example 18. We discuss the case $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ in which (3.5) corresponds to the primal update (U^k_P). Lemma 17 states that there exists an improvement direction according to (3.4) if and only if $(\boldsymbol{x}^k, 0)$ is not an optimal solution of (3.5). In that case, we obtain $t^{k+1} > 0$ and $\boldsymbol{x}^{k+1} \neq \boldsymbol{x}^k$. The specific improvement direction corresponding to the primal update is $\boldsymbol{d}^{k+1} = (\boldsymbol{x}^{k+1} - \boldsymbol{x}^k)/(\delta^k - \delta^{k+1})$ and each convex combination $\boldsymbol{x}^k + t\boldsymbol{d}^{k+1}$ for $0 \leq t \leq t^{k+1}$ induces an optimal pair for $(\mathrm{P}_{\delta^k - t})$. For $t > t^{k+1}$ we do not obtain further optimal pairs, since this would contradict the optimality of $(\boldsymbol{x}^{k+1}, t^{k+1})$ in (3.5).

So far, our method does not use explicit improvement directions. In fact, the updates (U^k_D) and (U^k_P) cannot simply be replaced by a search for improvement directions because the systems (3.2) and (3.4) do not necessarily have unique solutions. Nevertheless, improvement directions play an important role in the analysis of our method. The following lemma establishes a connection between the existence of dual and primal improvement directions.

Lemma 19. *Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an optimal pair for $(P_{\hat{\delta}})$ for some $\hat{\delta} > 0$, and let $\boldsymbol{\psi} \in \mathbb{R}^m$. Then, one and only one of the systems*

$$\begin{aligned} \boldsymbol{\psi}^\top \boldsymbol{e} &< 0 \\ \boldsymbol{A}_S^\top \boldsymbol{e} &= \boldsymbol{0} \\ \boldsymbol{A}_{\Sigma\setminus S}^\top \hat{\boldsymbol{y}} \odot \boldsymbol{A}_{\Sigma\setminus S}^\top \boldsymbol{e} &\leq \boldsymbol{0} \\ -\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{e}_{W\setminus\Omega} &\leq \boldsymbol{0} \\ \boldsymbol{e}_{W^c} &= \boldsymbol{0} \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} \boldsymbol{A}^\Omega \boldsymbol{d} &= \boldsymbol{\psi}_\Omega \\ \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}^{W\setminus\Omega}\boldsymbol{d} &\leq \operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{\psi}_{W\setminus\Omega} \\ \boldsymbol{A}_{\Sigma\setminus S}^\top \hat{\boldsymbol{y}} \odot \boldsymbol{d}_{\Sigma\setminus S} &\leq \boldsymbol{0} \\ \boldsymbol{d}_{\Sigma^c} &= \boldsymbol{0} \end{aligned} \tag{3.7}$$

has a solution.

Proof. With $\boldsymbol{T}_1 \coloneqq \operatorname{Diag}(\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}} - \boldsymbol{b}_{W\setminus\Omega}))$ and $\boldsymbol{T}_2 \coloneqq \operatorname{Diag}(\boldsymbol{A}_{\Sigma\setminus S}^\top \hat{\boldsymbol{y}})$, both $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are symmetric and orthogonal, and we can rewrite the first system as

$$\begin{aligned} \boldsymbol{\psi}_{W\setminus\Omega}^\top \boldsymbol{e}_{W\setminus\Omega} &+ & \boldsymbol{\psi}_\Omega^\top \boldsymbol{e}_\Omega &< 0 \\ (\boldsymbol{A}_S^{W\setminus\Omega})^\top \boldsymbol{e}_{W\setminus\Omega} &+ & (\boldsymbol{A}_S^\Omega)^\top \boldsymbol{e}_\Omega &= \boldsymbol{0} \\ \boldsymbol{T}_2(\boldsymbol{A}_{\Sigma\setminus S}^{W\setminus\Omega})^\top \boldsymbol{e}_{W\setminus\Omega} &+ & \boldsymbol{T}_2(\boldsymbol{A}_{\Sigma\setminus S}^\Omega)^\top \boldsymbol{e}_\Omega &\leq \boldsymbol{0} \\ -\boldsymbol{T}_1 \boldsymbol{e}_{W\setminus\Omega} & & &\leq \boldsymbol{0}. \end{aligned}$$

We substitute $\tilde{\boldsymbol{e}}_{W\setminus\Omega} \coloneqq \boldsymbol{T}_1 \boldsymbol{e}_{W\setminus\Omega}$ and observe that the system has a solution if and only if the system

$$\begin{aligned} \boldsymbol{\psi}_{W\setminus\Omega}^\top \boldsymbol{T}_1 \tilde{\boldsymbol{e}}_{W\setminus\Omega} &+ & \boldsymbol{\psi}_\Omega^\top \boldsymbol{e}_\Omega &< 0 \\ (\boldsymbol{A}_S^{W\setminus\Omega})^\top \boldsymbol{T}_1 \tilde{\boldsymbol{e}}_{W\setminus\Omega} &+ & (\boldsymbol{A}_S^\Omega)^\top \boldsymbol{e}_\Omega &= \boldsymbol{0} \\ -\boldsymbol{T}_2(\boldsymbol{A}_{\Sigma\setminus S}^{W\setminus\Omega})^\top \boldsymbol{T}_1 \tilde{\boldsymbol{e}}_{W\setminus\Omega} &- & \boldsymbol{T}_2(\boldsymbol{A}_{\Sigma\setminus S}^\Omega)^\top \boldsymbol{e}_\Omega &\geq \boldsymbol{0} \\ \tilde{\boldsymbol{e}}_{W\setminus\Omega} & & &\geq \boldsymbol{0} \end{aligned}$$

is feasible. By Farkas' Lemma [40, Corollary 7.1d], this system has a solution if and only if the associated alternative system

$$\begin{aligned} -\boldsymbol{A}_{\Sigma\setminus S}^\Omega \boldsymbol{T}_2 \tilde{\boldsymbol{d}}_{\Sigma\setminus S} &+ & \boldsymbol{A}_S^\Omega \boldsymbol{d}_S &= \boldsymbol{\psi}_\Omega \\ -\boldsymbol{T}_1 \boldsymbol{A}_{\Sigma\setminus S}^{W\setminus\Omega} \boldsymbol{T}_2 \tilde{\boldsymbol{d}}_{\Sigma\setminus S} &+ & \boldsymbol{T}_1 \boldsymbol{A}_S^{W\setminus\Omega} \boldsymbol{d}_S &\leq \boldsymbol{T}_1 \boldsymbol{\psi}_{W\setminus\Omega} \\ \tilde{\boldsymbol{d}}_{\Sigma\setminus S} & & &\geq \boldsymbol{0} \end{aligned}$$

is infeasible. Substituting $\boldsymbol{d}_{\Sigma\setminus S} \coloneqq -\boldsymbol{T}_2\tilde{\boldsymbol{d}}_{\Sigma\setminus S}$, we obtain that equivalently,

$$\begin{aligned} \boldsymbol{A}_\Sigma^\Omega \boldsymbol{d}_\Sigma &= \boldsymbol{\psi}_\Omega \\ \boldsymbol{T}_1\boldsymbol{A}_\Sigma^{W\setminus\Omega}\boldsymbol{d}_\Sigma &\leq \boldsymbol{T}_1\boldsymbol{\psi}_{W\setminus\Omega} \\ -\boldsymbol{T}_2\boldsymbol{d}_{\Sigma\setminus S} &\geq \boldsymbol{0} \end{aligned}$$

is infeasible. □

We see that the systems (3.2) and (3.6) which correspond to dual improvement directions are equal for arbitrary $\boldsymbol{\psi} \in \mathbb{R}^m$. On the other hand, the systems (3.4) and (3.7) that correspond to primal improvement directions are equal if and only if $\boldsymbol{\psi}_W = -\text{sign}(\boldsymbol{A}^W\hat{\boldsymbol{x}} - \boldsymbol{b}_W)$, and this is exactly key to the choice of $\boldsymbol{\psi}$. The proposed choice establishes a direct connection between the existence of dual and primal improvement directions via Lemma 19. Using Lemmas 15 and 17, this connection can be extended to the updates (U_D^k) and (U_P^k), as we will see in the following theorem. Note that $\boldsymbol{\psi}_{W^c}$ does not play any role as it does not appear in (3.7) and because of the equation $\boldsymbol{e}_{W^c} = \boldsymbol{0}$ in (3.2) and (3.6). However, we set $\boldsymbol{\psi} = -\text{sign}(\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b})$ for reasons of uniformity.

Theorem 20. *Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an optimal pair for $(P_{\hat{\delta}})$ for some $\hat{\delta} > \delta$. Moreover, let $\boldsymbol{\psi} = -\text{sign}(\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b})$. Then, the following four alternatives are equivalent.*

(I) *The system (3.2) is feasible.*

(II) *The vector $\hat{\boldsymbol{y}}$ is not an optimal solution of (3.3).*

(III) *The system (3.4) is infeasible.*

(IV) *The tuple $(\hat{\boldsymbol{x}}, 0)$ is an optimal solution of (3.5).*

Proof. Lemma 15 shows that alternatives (I) and (II) are equivalent, Lemma 17 shows that alternatives (III) and (IV) are equivalent and Lemma 19 shows that alternatives (I) and (III) are equivalent. □

3.4 ℓ_1-HOUDINI Algorithm and Finite Termination

Essentially, our method performs alternating dual and primal updates by solving the linear programs (U_D^k) and (U_P^k), respectively. As mentioned above, we assume that an optimal pair $(\boldsymbol{x}^0, \boldsymbol{y}^0)$ for some $\delta^k > \delta$ is known a priori. To that end, we make the simple observation that $\boldsymbol{x}^0 \coloneqq \boldsymbol{0}$ and $\boldsymbol{y}^0 \coloneqq \boldsymbol{0}$ form an optimal pair for (P_{δ^0}) whenever $\delta^0 \geq \|\boldsymbol{b}\|_\infty$. This follows directly if we plug $\boldsymbol{x}^0$, $\boldsymbol{y}^0$ and δ^0 into (3.1). Accordingly, we initialize $\delta^0 \coloneqq \|\boldsymbol{b}\|_\infty$ and proceed with the first dual update. Algorithm 1 illustrates the entire scheme. Note that we do not initialize $\boldsymbol{y}^0$ because we do not need it explicitly for the first dual update.

Input: $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $0 \leq \delta < \|\boldsymbol{b}\|_\infty$
Output: solution $\boldsymbol{x}^\star$ to problem (P_δ)

// Initialization:
$\delta^0 \leftarrow \|\boldsymbol{b}\|_\infty$
$\boldsymbol{x}^0 \leftarrow \boldsymbol{0}$
$S_0 \leftarrow \emptyset$
$W_0 \leftarrow \{i : |\boldsymbol{b}_i| = \delta^0\}$
$k \leftarrow 0$
repeat
  // Dual update:
  $\boldsymbol{y}^{k+1} \leftarrow$ solution of (U_D^k) with $\boldsymbol{\psi} = -\mathrm{sign}(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b})$
  $\Omega_{k+1} \leftarrow \left\{i : y_i^{k+1} \neq 0\right\}$
  $\Sigma_{k+1} \leftarrow \left\{j : |\boldsymbol{A}_j^\top \boldsymbol{y}^{k+1}| = 1\right\}$
  // Primal update:
  $(\boldsymbol{x}^{k+1}, t^{k+1}) \leftarrow$ solution of (U_P^k)
  $\delta^{k+1} \leftarrow \delta^k - t^{k+1}$
  $S_{k+1} \leftarrow \left\{j : x_j^{k+1} \neq 0\right\}$
  $W_{k+1} \leftarrow \left\{i : |\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i| = \delta^{k+1}\right\}$
  $k \leftarrow k + 1$
until $\delta^k = \delta$
return $\boldsymbol{x}^\star = \boldsymbol{x}^k$

Algorithm 1: ℓ_1-Houdini

Lemma 21. *In each two consecutive iterations, Algorithm 1 produces iterates $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$ and $\boldsymbol{x}^{k+1} \neq \boldsymbol{x}^k$. In particular, it holds that $t^{k+1} > 0$ in each iteration.*

Proof. In the beginning, we have $\boldsymbol{x}^0 = \boldsymbol{0}$ and determine $\boldsymbol{y}^1$ by solving (U_D^0), which corresponds to (3.3) with $\hat{\boldsymbol{x}} = \boldsymbol{x}^0$. Theorem 20 states that $(\boldsymbol{x}^0, 0)$ is not an optimal solution of (3.5) with $\hat{\boldsymbol{y}} = \boldsymbol{y}^1$ and $\hat{\delta} = \delta^0$, which is exactly the problem (U_P^0) that is solved in the first primal update. By construction, $(\boldsymbol{x}^1, t^1)$ is a solution of (U_P^0). Hence, it holds that $t^1 > 0$ and $\boldsymbol{x}^1 \neq \boldsymbol{x}^0$.

If $k \geq 1$, then $(\boldsymbol{x}^k, t^k)$ is, by construction, an optimal solution of (U_P^{k-1}), which corresponds to (3.5) with $\hat{\boldsymbol{y}} = \boldsymbol{y}^k$ and $\hat{\delta} = \delta^{k-1}$. In this problem, substituting t with $\tilde{t} \coloneqq t - t^k$ naturally leaves the optimal value $t = t^k$ unchanged. Hence, it holds that $\tilde{t} = 0$. On the other hand, the transformed problem is exactly (3.5) with $\hat{\boldsymbol{y}} = \boldsymbol{y}^k$ and $\hat{\delta} = \delta^{k-1} - t^k = \delta^k$, and $(\boldsymbol{x}^k, 0)$ is an optimal solution of this problem. As $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal pair for (P_{δ^k}), Theorem 20 states that $\boldsymbol{y}^k$ is not an optimal solution of the problem (3.3) with $\hat{\boldsymbol{x}} = \boldsymbol{x}^k$, which corresponds to (U_D^k). By construction, $\boldsymbol{y}^{k+1}$ is an optimal solution of this problem and it follows that $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$. In turn, because $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ is still an optimal pair for (P_{δ^k}), Theorem 20 states that $(\boldsymbol{x}^k, 0)$ is not an

optimal solution of (3.5) with $\hat{\boldsymbol{y}} = \boldsymbol{y}^{k+1}$ and $\hat{\delta} = \delta^k$, which is exactly (U_P^k). Hence, it holds that $t^{k+1} > 0$ and $\boldsymbol{x}^{k+1} \neq \boldsymbol{x}^k$. □

Clearly, Lemma 21 does not yet prove the convergence of Algorithm 1. Nevertheless, each iteration contributes at least a small portion towards a solution of (P_δ).

Theorem 22. *Algorithm 1 terminates after a finite number of iterations and returns an optimal solution of (P_δ).*

Proof. The number of possible support sets S_k, active sets W_k, associated sign patterns and combinations thereof is finite. Suppose that for $k < \ell$, Algorithm 1 yields $S \coloneqq S_k = S_\ell$, $W \coloneqq W_k = W_\ell$, $\operatorname{sign}(\boldsymbol{x}_S^k) = \operatorname{sign}(\boldsymbol{x}_S^\ell)$ and $\operatorname{sign}(\boldsymbol{A}^W\boldsymbol{x}^k - \boldsymbol{b}_W) = \operatorname{sign}(\boldsymbol{A}^W\boldsymbol{x}^\ell - \boldsymbol{b}_W)$. Consequently, (U_D^k) and (U_D^ℓ) coincide and we obtain that $\boldsymbol{y}^{k+1} = \boldsymbol{y}^{\ell+1}$. It follows that (U_P^k) and (U_P^ℓ) are equal except that, due to $k < \ell$ and Lemma 21, we have $\delta^k > \delta^\ell$. Since δ^k and δ^ℓ are constants in the respective problems, it is equivalent to rewrite the objective functions as $t - \delta^k$ and $t - \delta^\ell$, respectively. The substitutions $\tilde{\delta} \coloneqq \delta^k - t$ and $\tilde{\delta} \coloneqq \delta^\ell - t$, respectively, then reveal that (U_P^k) and (U_P^ℓ) indeed have an identical reformulation. Hence, we obtain the same optimal value for $\tilde{\delta}$ in both cases, which shows that $\delta^{k+1} = \delta^k - t = \tilde{\delta} = \delta^\ell - t = \delta^{\ell+1}$. Since $k < \ell$, this contradicts Lemma 21. Thus, Algorithm 1 terminates after a finite number of iterations with an optimal solution. □

Note that we have assumed throughout that the feasible set of the problem (P_δ) is non-empty. If this is not the case, ℓ_1-Houdini can easily be modified in order to return an optimal solution associated with the smallest homotopy parameter $\delta_{\min}$ for which the feasible set of $(\mathrm{P}_{\delta_{\min}})$ is non-empty. To that end, we can simply check in each iteration whether the step size t^{k+1} is strictly positive. In case $t^{k+1} = 0$, it holds that $\delta_{\min} = \delta^k$ and $\boldsymbol{x}^k$ is an associated optimal solution.

3.5 Analysis of the Solution Path

While the finite termination of Algorithm 1 is now evident, we concentrate our attention on the set of all optimal solutions for $\delta \in [0, \infty)$. As a central result, we show that Algorithm 1 does not only generate optimal solutions $\boldsymbol{x}^k$ associated with the values δ^k, but implicitly also optimal solutions $\boldsymbol{x}^\star(\delta)$ for each possible value δ.

Definition 23. Let $\boldsymbol{x}^0, \ldots, \boldsymbol{x}^K$ denote the iterates that result from applying Algorithm 1 to the problem (P_0) and let $\delta^0 > \cdots > \delta^K = 0$ denote the corresponding values of the homotopy parameter. We define $\boldsymbol{x}^\star(\delta) \coloneqq \mathbf{0}$ for $\delta \geq \delta^0$ and

$$\boldsymbol{x}^\star(\delta) \coloneqq \boldsymbol{x}^k + (\delta^k - \delta)\frac{\boldsymbol{x}^{k+1} - \boldsymbol{x}^k}{\delta^k - \delta^{k+1}}$$

for $\delta^{k+1} \leq \delta < \delta^k$. We call $\boldsymbol{x}^\star : [0, \infty) \to \mathbb{R}^n$ the *primal solution mapping* and refer to

$$\mathcal{P} \coloneqq \{\boldsymbol{x}^\star(\delta) : \delta \in [0, \infty)\}$$

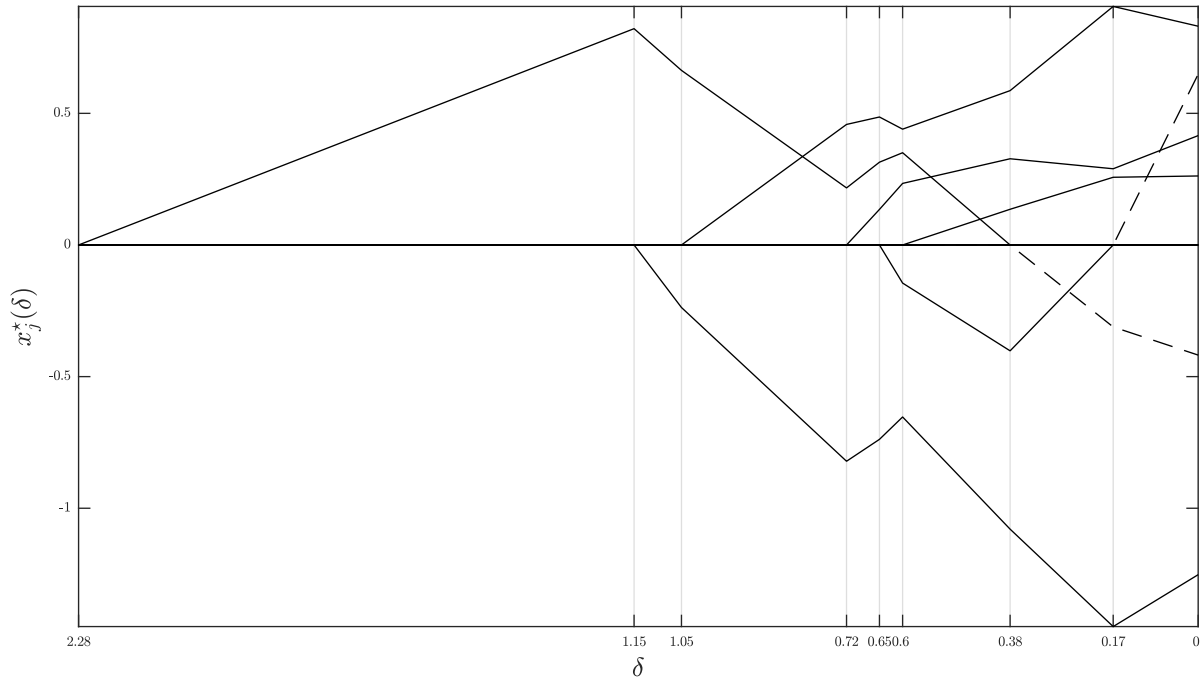

Figure 3.1: Example of a primal solution path generated by ℓ_1-HOUDINI.

as the *primal solution path*. Moreover,

$$S_\delta \coloneqq \{j : x_j^\star(\delta) \neq 0\} \quad \text{and} \quad W_\delta \coloneqq \{i : |\boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i| = \delta\}$$

denote the support and the active set of $\boldsymbol{x}^\star(\delta)$.

As we already discussed in Example 18, the vector $\boldsymbol{d}^{k+1} \coloneqq (\boldsymbol{x}^{k+1} - \boldsymbol{x}^k)/(\delta^k - \delta^{k+1})$ is the improvement direction that corresponds to the update (U_P^k). Therefore, the points $\boldsymbol{x}^k$ can be interpreted as *kinks* where the primal solution path changes direction. Likewise, each set $\mathcal{P}_{k+1} \coloneqq \{\boldsymbol{x}^\star(\delta) : \delta \in [\delta^{k+1}, \delta^k]\}$ forms a linear *segment* of the primal solution path.

Figure 3.1 shows an exemplary run of ℓ_1-HOUDINI with $\boldsymbol{A} \in \mathbb{R}^{6\times 12}$ and $\boldsymbol{b} \in \mathbb{R}^6$ generated at random and $\delta = 0$, where the algorithm needed 9 iterations to solve the problem. Horizontal labels display the value of the homotopy parameter δ^k after each iteration. Each vertical line marks a kink and separates two linear segments of the primal solution path. The plots represent the primal solution path on the level of the components $x_j^\star(\delta)$ for $j = 1, \ldots, 12$. The dashed lines indicate that one variable leaves the support and another one becomes non-zero at the respective kinks. We continue by proving some essential properties of the primal solution mapping.

Lemma 24. *The primal solution mapping $\boldsymbol{x}^\star : [0, \infty) \to \mathbb{R}^n$ is continuous piecewise linear. In particular, it is linear on each interval $[\delta^{k+1}, \delta^k]$ and constant on $[\delta^0, \infty)$.*

Proof. The linearity on each interval follows directly from Definition 23. Further, it holds for each $k \in \{0, \dots, K-1\}$ that

$$\lim_{\delta \nearrow \delta^k} \boldsymbol{x}^\star(\delta) = \boldsymbol{x}^k = \lim_{\delta \searrow \delta^k} \boldsymbol{x}^\star(\delta)$$

and hence, $\boldsymbol{x}^\star$ is continuous at each intersection point of two intervals. □

Lemma 25. *For $\delta \in [0, \infty)$, it holds that $\boldsymbol{x}^\star(\delta)$ is an optimal solution of (P_δ).*

Proof. For $\delta \geq \|\boldsymbol{b}\|_\infty = \delta^0$, the all-zero vector is an optimal solution of (P_δ) and hence, the statement is true. So let $\delta^{k+1} \leq \delta \leq \delta^k$ for some k. By construction, $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ and $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ form optimal pairs for (P_{δ^k}) and ($\mathrm{P}_{\delta^{k+1}}$), respectively. By Lemma 21, it further holds that $\delta^{k+1} < \delta^k$. Therewith, it follows from Lemma 17 that $(\boldsymbol{x}^k + t\boldsymbol{d}^{k+1}, \boldsymbol{y}^{k+1})$ is an optimal pair for ($\mathrm{P}_{\delta^k - t}$) for each $0 \leq t \leq \delta^k - \delta^{k+1}$. If we set $\delta \coloneqq \delta^k - t$, we see that $(\boldsymbol{x}^k + (\delta^k - \delta)\boldsymbol{d}^{k+1}, \boldsymbol{y}^{k+1})$ is an optimal pair for (P_δ) for all $\delta^{k+1} \leq \delta \leq \delta^k$. □

In the following, we study the properties of dual solutions that are associated with fixed points or segments along the primal solution path. Besides the particular dual solutions determined by Algorithm 1, our analysis explicitly includes all other existing dual solutions. On the one hand, this enables us to illustrate the set of all dual solutions related to the primal solution path. On the other hand, we obtain a useful tool in order to estimate the number of iterations in Algorithm 1.

Lemma 26. *Let $(\hat{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ be an optimal pair for ($P_{\hat{\delta}}$) for some $\hat{\delta} \geq 0$. Then, it holds that*

$$\operatorname{sign}(\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b})^\top \tilde{\boldsymbol{y}} = \|\tilde{\boldsymbol{y}}\|_1.$$

Proof. Let $\tilde{\Omega}$ denote the support of $\tilde{\boldsymbol{y}}$. Using the conditions ($\mathrm{C}_{\hat{\delta}}$) and $\tilde{\Omega} \subseteq W$, we obtain

$$\begin{aligned}
\operatorname{sign}(\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b})^\top \tilde{\boldsymbol{y}} &= \operatorname{sign}(\boldsymbol{A}^{\tilde{\Omega}}\hat{\boldsymbol{x}} - \boldsymbol{b}_{\tilde{\Omega}})^\top \tilde{\boldsymbol{y}}_{\tilde{\Omega}} \\
&= \frac{1}{\hat{\delta}} (\boldsymbol{A}^{\tilde{\Omega}}\hat{\boldsymbol{x}} - \boldsymbol{b}_{\tilde{\Omega}})^\top \tilde{\boldsymbol{y}}_{\tilde{\Omega}} \\
&= \frac{1}{\hat{\delta}} \hat{\delta} \operatorname{sign}(\tilde{\boldsymbol{y}}_{\tilde{\Omega}})^\top \tilde{\boldsymbol{y}}_{\tilde{\Omega}} = \|\tilde{\boldsymbol{y}}\|_1.
\end{aligned}$$

□

As a consequence of Lemma 26, we see that in each dual update (U_D^k) with the objective function $\boldsymbol{\psi} = -\operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b})$, we actually pick one dual solution $\boldsymbol{y}^{k+1}$ that has maximal ℓ_1-norm among all dual solutions that form an optimal pair with $\boldsymbol{x}^k$. Our next step is to show that the norm of all dual solutions at a kink lies in a certain interval, while the norm of all dual solutions associated with the relative interior of a linear segment is constant.

Lemma 27. *Let $(\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}})$ be an optimal pair for (P_δ) for some $\delta \in (\delta^{k+1}, \delta^k]$. Then, it holds that $\|\tilde{\boldsymbol{y}}\|_1 \leq \|\boldsymbol{y}^{k+1}\|_1$.*

Proof. In view of Lemma 26, we conclude that by solving (3.3) with $\hat{\boldsymbol{x}} = \boldsymbol{x}^k$, we implicitly determine a dual solution $\boldsymbol{y}^{k+1}$ with maximal ℓ_1-norm among all dual solutions that form an optimal pair together with $\boldsymbol{x}^k$. Hence, the statement holds for $\delta = \delta^k$. Moreover, Lemma 17 with $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ and $\tilde{\boldsymbol{x}} = \boldsymbol{x}^{k+1}$ states that $(\boldsymbol{x}^\star(\delta), \boldsymbol{y}^{k+1})$ is also an optimal pair for *each* $\delta \in (\delta^{k+1}, \delta^k)$. With $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^\star(\delta), \boldsymbol{y}^{k+1})$, $\boldsymbol{d}^{k+1}$ is a primal improvement direction according to (3.4) for each $\delta \in (\delta^{k+1}, \delta^k)$. Thus, Theorem 20 and Lemma 26 imply that $\boldsymbol{y}^{k+1}$ is a dual solution with maximal ℓ_1-norm among all dual solutions that form an optimal pair with *some* $\boldsymbol{x}^\star(\delta)$ for $\delta \in (\delta^{k+1}, \delta^k)$. □

Lemma 28. *Let $\delta \in (\delta^{k+1}, \delta^k)$. Then, the associated primal improvement direction satisfies*

$$\boldsymbol{d}^{k+1}_{S^c_\delta} = \boldsymbol{0} \quad \textit{and} \quad \operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta}) \odot \boldsymbol{A}^{W_\delta}\boldsymbol{d}^{k+1} = -\mathbb{1}.$$

Moreover, it holds that $S_\delta = S_k \cup S_{k+1}$ and $W_\delta \subseteq W_k \cap W_{k+1}$, and the sets S_δ and W_δ are invariant on (δ^{k+1}, δ^k).

Proof. Due to the fact that $\delta \in (\delta^{k+1}, \delta^k)$, $\boldsymbol{x}^\star(\delta)$ is a non-trivial convex combination of $\boldsymbol{x}^k$ and $\boldsymbol{x}^{k+1}$. Moreover, $\boldsymbol{x}^k$ and $\boldsymbol{x}^{k+1}$ cannot have opposing sign patterns because both form an optimal pair with the same dual solution $\boldsymbol{y}^{k+1}$. We conclude that $S_\delta = S_k \cup S_{k+1}$ and that S_δ is invariant on (δ^{k+1}, δ^k). It follows that $\boldsymbol{x}^k_{S^c_\delta} = \boldsymbol{x}^{k+1}_{S^c_\delta} = \boldsymbol{0}$ and hence, $\boldsymbol{d}^{k+1}_{S^c_\delta} = \boldsymbol{0}$.

For the second part, we first show that $W_\delta \subseteq W_k$. To this end, suppose that there exists an $i \in W_\delta \setminus W_k$ with $\boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i = \delta$. This implies $\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1} > -1$ and

$$\delta = \boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i = \boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i + (\delta^k - \delta)\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1}.$$

Therewith, it follows that

$$\begin{aligned}
\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i &= \boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta^{k+1}) - b_i \\
&= \boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i + (\delta^k - \delta^{k+1})\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1} \\
&= \delta + (\delta - \delta^{k+1})\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1} \\
&> \delta^{k+1}.
\end{aligned}$$

If $\boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i = -\delta$, we obtain analogously that $\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i < -\delta^{k+1}$. Either way, $\boldsymbol{x}^{k+1}$ is infeasible for $(\mathrm{P}_{\delta^{k+1}})$. As $\boldsymbol{x}^{k+1}$ is, by construction, an optimal soution of $(\mathrm{P}_{\delta^{k+1}})$, this is a contradiction and it follows that $W_\delta \subseteq W_k$. A similar reasoning can be used to show that $W_\delta \subseteq W_{k+1}$. We assume that there exists an $i \in W_\delta \setminus W_{k+1}$ with $\boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i = \delta$, observe that then $\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1} < -1$ and $\delta = \boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i - (\delta - \delta^{k+1})\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1}$, and conclude that $\boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i > \delta^k$. Analogously, we argue that $\boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i < -\delta^k$ in case $\boldsymbol{a}_i^\top \boldsymbol{x}^\star(\delta) - b_i = -\delta$. All in all, it follows that $W_\delta \subseteq W_k \cap W_{k+1}$.

As $W_\delta \subseteq W_k \cap W_{k+1}$ for all $\delta \in (\delta^{k+1}, \delta^k)$, the continuity of the primal solution mapping shows that $\operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta}) = \operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^k - \boldsymbol{b}_{W_\delta}) = \operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^{k+1} - \boldsymbol{b}_{W_\delta})$ holds for

all $\delta \in (\delta^{k+1}, \delta^k)$. In combination with the optimality conditions for (P_{δ^k}) and $(\mathrm{P}_{\delta^{k+1}})$, we further obtain that

$$\begin{aligned}\operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta}) \odot (\boldsymbol{A}^{W_\delta}\boldsymbol{x}^k - \boldsymbol{b}_{W_\delta}) &= \delta^k \mathbb{1} \text{ and} \\ \operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta}) \odot (\boldsymbol{A}^{W_\delta}\boldsymbol{x}^{k+1} - \boldsymbol{b}_{W_\delta}) &= \delta^{k+1} \mathbb{1}.\end{aligned}$$

Subtracting the first equality from the second and dividing by $\delta^k - \delta^{k+1}$ shows that $\operatorname{sign}(\boldsymbol{A}^{W_\delta}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta}) \odot \boldsymbol{A}^{W_\delta}\boldsymbol{d}^{k+1} = -\mathbb{1}$. Hence, W_δ is invariant on (δ^{k+1}, δ^k). □

In view of Lemma 28, notice that not necessarily $W_k \cap W_{k+1} \subseteq W_\delta$. In fact, it may happen that there exists an $i \in W_k \cap W_{k+1}$ such that

$$\operatorname{sign}(\boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i) = -\operatorname{sign}(\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i), \tag{3.8}$$

i.e., the i-th constraint is active at both $\boldsymbol{x}^k$ and $\boldsymbol{x}^{k+1}$ but with opposing signs. This implies $|\boldsymbol{a}_i^\top \boldsymbol{d}^{k+1}| > 1$ and consequently $i \notin W_\delta$.

Lemma 29. *Let $(\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}})$ be an optimal pair for (P_δ) for some $\delta \in (\delta^{k+1}, \delta^k)$. Then, it holds that $\|\tilde{\boldsymbol{y}}\|_1 = \|\boldsymbol{y}^{k+1}\|_1$.*

Proof. We consider an optimal pair $(\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}})$ for (P_δ) for some $\delta \in (\delta^{k+1}, \delta^k)$. Lemma 27 already shows that $\|\tilde{\boldsymbol{y}}\|_1 \leq \|\boldsymbol{y}^{k+1}\|_1$, so let us assume that $\|\tilde{\boldsymbol{y}}\|_1 < \|\boldsymbol{y}^{k+1}\|_1$ and derive a contradiction. By Lemma 26, our assumption is equivalent to

$$\operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^\star(\delta) - \boldsymbol{b})^\top \tilde{\boldsymbol{y}} < \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^\star(\delta) - \boldsymbol{b})^\top \boldsymbol{y}^{k+1}. \tag{3.9}$$

We apply Lemma 15 with $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^\star(\delta), \boldsymbol{y}^{k+1})$ and $\boldsymbol{\psi} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^\star(\delta) - \boldsymbol{b})$. As $(\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}})$ is an optimal pair as well, both $\boldsymbol{y}^{k+1}$ and $\tilde{\boldsymbol{y}}$ satisfy (3.3b)–(3.3e). Because of (3.9), we see that $\boldsymbol{y}^{k+1}$ is not an optimal solution of (3.3) and that $\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{y}} - \boldsymbol{y}^{k+1}$ is feasible for (3.2). Now, we apply Lemma 19 which shows that there exists no solution of the system

$$\begin{aligned} \boldsymbol{A}^{\Omega_{k+1}}\boldsymbol{d} &= \operatorname{sign}(\boldsymbol{y}^{k+1}_{\Omega_{k+1}}) \\ \operatorname{sign}(\boldsymbol{A}^{W_\delta \setminus \Omega_{k+1}}\boldsymbol{x}^\star(\delta) - \boldsymbol{b}_{W_\delta \setminus \Omega_{k+1}}) \odot \boldsymbol{A}^{W_\delta \setminus \Omega_{k+1}}\boldsymbol{d} &\leq \mathbb{1} \\ \boldsymbol{A}^\top_{\Sigma_{k+1} \setminus S_\delta} \boldsymbol{y}^{k+1} \odot \boldsymbol{d}_{\Sigma_{k+1} \setminus S_\delta} &\leq \boldsymbol{0} \\ \boldsymbol{d}_{\Sigma^c_{k+1}} &= \boldsymbol{0}. \end{aligned} \tag{3.10}$$

At the same time, Lemma 28 shows that $-\boldsymbol{d}^{k+1}$ satisfies the second and third conditions in (3.10), and Lemma 17 with $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ and $\tilde{\boldsymbol{x}} = \boldsymbol{x}^{k+1}$ shows that $-\boldsymbol{d}^{k+1}$ satisfies the first and fourth conditions in (3.10) as well. Thus, we have found a contradiction and conclude that indeed $\|\tilde{\boldsymbol{y}}\|_1 = \|\boldsymbol{y}^{k+1}\|_1$. □

Lemma 30. *Let $(\boldsymbol{x}^{k+1}, \tilde{\boldsymbol{y}})$ be an optimal pair for $(P_{\delta^{k+1}})$. Then, it holds that $\|\tilde{\boldsymbol{y}}\|_1 \geq \|\boldsymbol{y}^{k+1}\|_1$.*

Proof. Suppose that $(\boldsymbol{x}^{k+1}, \tilde{\boldsymbol{y}})$ is an optimal pair for $(\mathrm{P}_{\delta^{k+1}})$ and $\|\tilde{\boldsymbol{y}}\|_1 < \|\boldsymbol{y}^{k+1}\|_1$. We proceed analogously to the proof of Lemma 29, this time with $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ and $\boldsymbol{\psi} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}^{k+1} - \boldsymbol{b})$. In this way, we obtain that there exists no solution of the system

$$\begin{aligned} \boldsymbol{A}^{\Omega_{k+1}}\boldsymbol{d} &= \operatorname{sign}(\boldsymbol{y}^{k+1}_{\Omega_{k+1}}) \\ \operatorname{sign}(\boldsymbol{A}^{W_{k+1}\setminus\Omega_{k+1}}\boldsymbol{x}^{k+1} - \boldsymbol{b}_{W_{k+1}\setminus\Omega_{k+1}}) \odot \boldsymbol{A}^{W_{k+1}\setminus\Omega_{k+1}}\boldsymbol{d} &\leq \mathbb{1} \\ \boldsymbol{A}^{\top}_{\Sigma_{k+1}\setminus S_{k+1}}\boldsymbol{y}^{k+1} \odot \boldsymbol{d}_{\Sigma_{k+1}\setminus S_{k+1}} &\leq \boldsymbol{0} \\ \boldsymbol{d}_{\Sigma^{\mathrm{c}}_{k+1}} &= \boldsymbol{0}. \end{aligned} \tag{3.11}$$

Again, Lemma 17 with $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = (\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ and $\tilde{\boldsymbol{x}} = \boldsymbol{x}^{k+1}$ shows that $-\boldsymbol{d}^{k+1}$ satisfies the first and fourth conditions in (3.11). Now, consider the second condition. If $i \in W_\delta \setminus \Omega_{k+1}$ for $\delta \in (\delta^{k+1}, \delta^k)$, then the inequality holds by Lemma 28. For each $i \in W_{k+1} \setminus W_\delta$, it holds that $\operatorname{sign}(\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i) \cdot \boldsymbol{a}_i^\top \boldsymbol{d}^{k+1} > -1$. Consequently, the second condition is valid for $-\boldsymbol{d}^{k+1}$. Finally, we show that the third condition is valid for $-\boldsymbol{d}^{k+1}$. If $j \notin S_\delta$ for $\delta \in (\delta^{k+1}, \delta^k)$, then we have $d_j^{k+1} = 0$ by Lemma 28. For $j \in S_\delta \setminus S_{k+1}$, we obtain $j \in S_k$ by Lemma 28. The optimality of $\boldsymbol{x}^k$ for (P_{δ^k}) now implies that $\boldsymbol{A}_j^\top \boldsymbol{y}^{k+1} \cdot x_j^k < 0$. Since $0 = \boldsymbol{x}_j^{k+1} = \boldsymbol{x}_j^k + (\delta^k - \delta^{k+1})\boldsymbol{d}_j^{k+1}$, we see that $\boldsymbol{A}_j^\top \boldsymbol{y}^{k+1} \cdot d_j^{k+1} > 0$. Hence, $-\boldsymbol{d}^{k+1}$ satisfies the third condition and is thus a solution of (3.11). □

Proposition 31. *Consider the multivalued mapping* $\mathcal{F} : [0, \infty) \rightrightarrows [0, \infty)$ *with*

$$\mathcal{F}(\delta) \coloneqq \{\|\tilde{\boldsymbol{y}}\|_1 : (\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}}) \text{ is an optimal pair}\}.$$

Then, it holds that

$$\mathcal{F}(\delta) = \begin{cases} [\|\boldsymbol{y}^k\|_1, \|\boldsymbol{y}^{k+1}\|_1], & \text{if } \delta = \delta^k \\ \{\|\boldsymbol{y}^{k+1}\|_1\}, & \text{if } \delta \in (\delta^{k+1}, \delta^k) \end{cases}$$

with $\|\boldsymbol{y}^k\|_1 < \|\boldsymbol{y}^{k+1}\|_1$.

Proof. For $\delta \in (\delta^{k+1}, \delta^k)$, Lemma 29 states that each dual solution inducing an optimal pair $(\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}})$ satisfies $\|\tilde{\boldsymbol{y}}\|_1 = \|\boldsymbol{y}^{k+1}\|_1$. It remains to show that the statement is true if $\delta = \delta^k$ for each $k \in \{0, \ldots, K\}$. Notice that, for ease of notation, we implicitly assumed that $\boldsymbol{y}^0 = \boldsymbol{0}$ and that $\boldsymbol{y}^{K+1}$ is determined according to (U_D^{K+1}). For each optimal pair $(\boldsymbol{x}^k, \tilde{\boldsymbol{y}})$, Lemma 27 states that $\|\tilde{\boldsymbol{y}}\|_1 \leq \|\boldsymbol{y}^{k+1}\|_1$ and by Lemma 30, it holds that $\|\tilde{\boldsymbol{y}}\|_1 \geq \|\boldsymbol{y}^k\|$. As both $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ and $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ are optimal pairs, Lemma 15 states that, for each $0 \leq s \leq 1$, $(\boldsymbol{x}^k, s\boldsymbol{y}^{k+1} + (1-s)\boldsymbol{y}^k)$ is also an optimal pair. Moreover, both $\boldsymbol{y}^k$ and $\boldsymbol{y}^{k+1}$ satisfy (3.3d) which means that they cannot have opposing sign patterns. Thus, we have $\{\|s\boldsymbol{y}^{k+1} + (1-s)\boldsymbol{y}^k\|_1 : 0 \leq s \leq 1\} = [\|\boldsymbol{y}^k\|_1, \|\boldsymbol{y}^{k+1}\|_1]$. Lemma 21 states that $\boldsymbol{y}^{k+1} \neq \boldsymbol{y}^k$. Analogous to the proof of Lemma 21, we obtain from Theorem 20 and Lemma 26 that $\|\boldsymbol{y}^k\|_1 < \|\boldsymbol{y}^{k+1}\|_1$. □

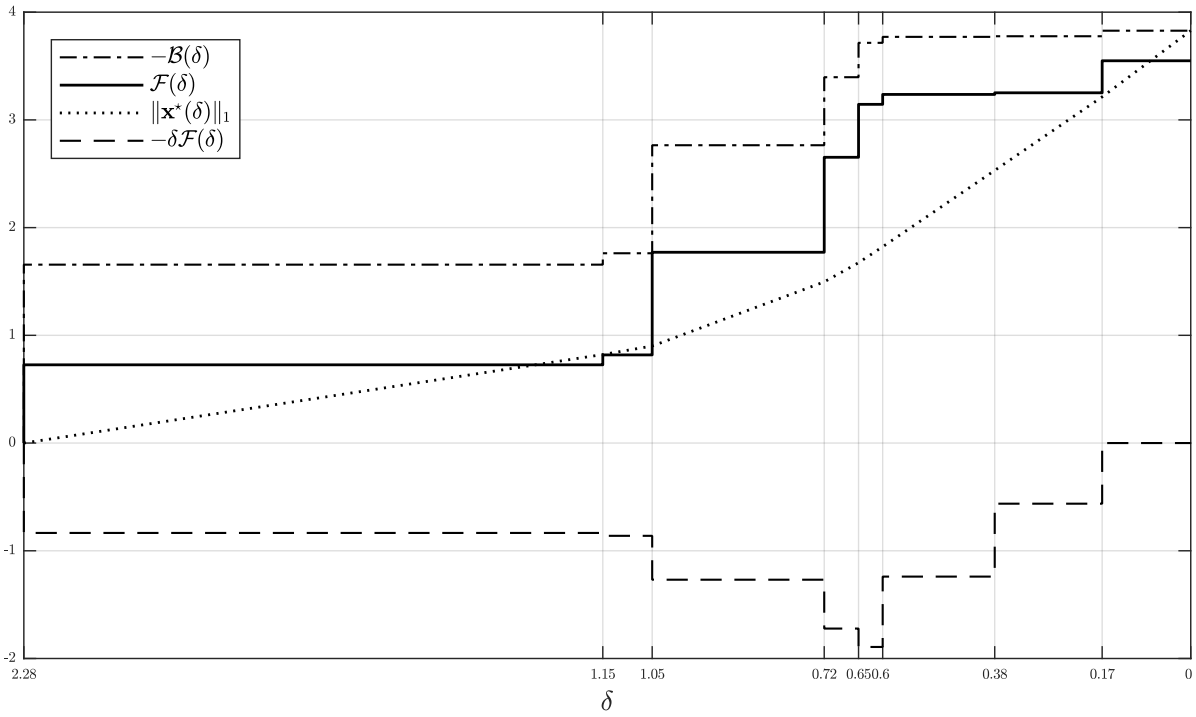

Figure 3.2: The mappings $\mathcal{F}(\delta)$ and $\mathcal{B}(\delta)$ in the context of Fenchel-Rockafellar duality.

Corollary 32. *Consider the multivalued mapping* $\mathcal{B} : [0, \infty) \rightrightarrows [0, \infty)$ *with*

$$\mathcal{B}(\delta) \coloneqq \{\boldsymbol{b}^\top \tilde{\boldsymbol{y}} : (\boldsymbol{x}^\star(\delta), \tilde{\boldsymbol{y}}) \text{ is an optimal pair}\}$$

Then, it holds that

$$\mathcal{B}(\delta) = \begin{cases} [\boldsymbol{b}^\top \boldsymbol{y}^{k+1}, \boldsymbol{b}^\top \boldsymbol{y}^k], & \text{if } \delta = \delta^k \\ \{\boldsymbol{b}^\top \boldsymbol{y}^{k+1}\}, & \text{if } \delta \in (\delta^{k+1}, \delta^k) \end{cases}$$

with $\boldsymbol{b}^\top \boldsymbol{y}^{k+1} < \boldsymbol{b}^\top \boldsymbol{y}^k \leq 0$.

Proof. By Fenchel-Rockafellar duality, it holds that $\|\boldsymbol{x}^\star(\delta)\|_1 = -\boldsymbol{b}^\top \tilde{\boldsymbol{y}} - \delta \|\tilde{\boldsymbol{y}}\|_1$, whenever $\boldsymbol{x}^\star(\delta)$ and $\tilde{\boldsymbol{y}}$ form an optimal pair. Hence, it holds that $\mathcal{B}(\delta) = -\|\boldsymbol{x}^\star(\delta)\|_1 - \delta\mathcal{F}(\delta)$. This already shows that the statement is true in case $\delta \in (\delta^{k+1}, \delta^k)$, where $\mathcal{F}(\delta) = \{\|\boldsymbol{y}^{k+1}\|_1\}$ is single-valued. In case $\delta = \delta^k$, we have

$$\begin{aligned} \mathcal{B}(\delta^k) &= -\|\boldsymbol{x}^\star(\delta^k)\|_1 - \delta^k \mathcal{F}(\delta^k) \\ &= -\|\boldsymbol{x}^\star(\delta^k)\|_1 - \delta^k [\|\boldsymbol{y}^k\|_1, \|\boldsymbol{y}^{k+1}\|_1] \\ &= [-\|\boldsymbol{x}^\star(\delta^k)\|_1 - \delta^k \|\boldsymbol{y}^{k+1}\|_1, -\|\boldsymbol{x}^\star(\delta^k)\|_1 - \delta^k \|\boldsymbol{y}^k\|_1] \\ &= [\boldsymbol{b}^\top \boldsymbol{y}^{k+1}, \boldsymbol{b}^\top \boldsymbol{y}^k]. \end{aligned}$$

□

Proposition 31 implies that $\mathcal{F}$ is monotonically decreasing in terms of monotonocity of multivalued mappings, i.e., it holds for $\delta, \delta' \in [0, \infty)$, $\alpha \in \mathcal{F}(\delta)$ and $\alpha' \in \mathcal{F}(\delta')$ that $(\delta - \delta') \cdot (\alpha - \alpha') \leq 0$. Analogously, it follows from Corollary 32 that $\mathcal{B}$ is monotonically increasing.

Figure 3.2 revisits the example of Figure 3.1. The bold line illustrates the mapping $\mathcal{F}(\delta)$ on the entire range $[\delta^0, \delta^K]$. Moreover, it compares the ℓ_1-norm of the primal solution mapping $\boldsymbol{x}^\star$ with $-\mathcal{B}$ and $-\delta\mathcal{F}$ in the context of strong duality. As mentioned above, it holds that $\boldsymbol{x}^\star(\delta) = -\boldsymbol{b}^\top \tilde{\boldsymbol{y}} - \delta \|\tilde{\boldsymbol{y}}\|_1$ for some dual solution $\tilde{\boldsymbol{y}}$. The example verifies that $\mathcal{F}$ and $-\mathcal{B}$ are monotonically decreasing, while we cannot expect $-\delta\mathcal{F}$ to be monotone.

3.6 Upper Complexity Bounds

Proposition 33. *The number of iterations K in Algorithm 1 satisfies $K \leq 3^{m+n}$.*

Proof. The statement was already shown implicitly in the proof of Theorem 22. However, we prove it here utilizing the mapping $\mathcal{F}$. Suppose that Algorithm 1 produces iterates $\boldsymbol{x}^k$ and $\boldsymbol{x}^\ell$ such that $S_k = S_\ell$, $W_k = W_\ell$, $\operatorname{sign}(\boldsymbol{x}^k_{S_k}) = \operatorname{sign}(\boldsymbol{x}^\ell_{S_\ell})$ and $\operatorname{sign}(\boldsymbol{A}^{W_k}\boldsymbol{x}^k - \boldsymbol{b}_{W_k}) = \operatorname{sign}(\boldsymbol{A}^{W_\ell}\boldsymbol{x}^\ell - \boldsymbol{b}_{W_\ell})$. Lemma 13 shows that the conditions (C^k_D) and (C^ℓ_D) are equal and hence, $(\boldsymbol{x}^k, \tilde{\boldsymbol{y}})$ is an optimal pair if and only if $(\boldsymbol{x}^\ell, \tilde{\boldsymbol{y}})$ is an optimal pair. It follows that $\mathcal{F}(\delta^k) = \mathcal{F}(\delta^\ell)$ and hence, $k = \ell$. Now, the statement follows because there are exactly 3^{m+n} different combinations of S_k, W_k and the associated sign patterns. □

Theorem 34. *The number of iterations K in Algorithm 1 satisfies $K \leq \frac{3^{m+n}+1}{2}$.*

Proof. Suppose that Algorithm 1 produces iterates $\boldsymbol{x}^k$ and $\boldsymbol{x}^\ell$ which satisfy $S_k = S_\ell$, $W_k = W_\ell$, $\operatorname{sign}(\boldsymbol{x}^k_{S_k}) = -\operatorname{sign}(\boldsymbol{x}^\ell_{S_\ell})$ and $\operatorname{sign}(\boldsymbol{A}^{W_k}\boldsymbol{x}^k - \boldsymbol{b}_{W_k}) = -\operatorname{sign}(\boldsymbol{A}^{W_\ell}\boldsymbol{x}^\ell - \boldsymbol{b}_{W_\ell})$. Lemma 13 shows that the conditions (C^k_D) and (C^ℓ_D) are equal up to the opposing sign patterns and hence, $(\boldsymbol{x}^k, \tilde{\boldsymbol{y}})$ is an optimal pair if and only if $(\boldsymbol{x}^\ell, -\tilde{\boldsymbol{y}})$ is an optimal pair. It follows that $\mathcal{F}(\delta^k) = \mathcal{F}(\delta^\ell)$ and hence, $k = \ell$. According to Proposition 33, the number of possible combinations after identification of opposing sign patterns is $(3^{m+n} - 1)/2 + 1 = (3^{m+n} + 1)/2$. □

3.7 Lower Complexity Bounds

Our goal in this section is to show that the worst case complexity of ℓ_1-Houdini is indeed exponential in the number of variables. More precisely, after we have shown in the previous section that the number of iterations in ℓ_1-Houdini can not exceed $(3^{m+n} + 1)/2$, we show now that there exist instances where ℓ_1-Houdini has to perform exactly $(3^n + 1)/2$ iterations in order to find an optimal solution. To that end, we adopt the strategy that was proposed in [31] to show that the number of linear segments in the regularization path of the Lasso (see [42]) is exponential in the number of variables in the worst case. In short, we proceed as follows: Building on $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$

and the related problem (P_δ), we design $\tilde{\boldsymbol{A}} \in \mathbb{R}^{(m+1)\times(n+1)}$ and $\tilde{\boldsymbol{b}} \in \mathbb{R}^{m+1}$ and consider the associated problem ($\tilde{\mathrm{P}}_\eta$). Afterwards, we show that an optimal pair for ($\tilde{\mathrm{P}}_\eta$) can be constructed in terms of a particular optimal pair for (P_δ), where the value of δ depends on η. In the final step, we proceed to the solution path of ($\tilde{\mathrm{P}}_0$) and conclude, given that the solution path of (P_0) has $K+1$ linear segments, that it has exactly $3(K+1)-1$ linear segments. Iteratively applying this strategy shows that ℓ_1-Houdini has to perform at least $(3^n+1)/2$ iterations in the worst case.

Proposition 35. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}$ such that the problem (P_δ) has a unique solution for each $\delta \geq 0$. Further, let $\boldsymbol{x}^0, \dots, \boldsymbol{x}^K$ be the primal iterates produced by ℓ_1-Houdini applied to the problem (P_0), let $\|\boldsymbol{b}\|_\infty = \delta^0 > \cdots > \delta^K = 0$ denote the corresponding values of the homotopy parameter and let*

$$\boldsymbol{\sigma}^0 \coloneqq \operatorname{sign}(\boldsymbol{x}^0), \quad \boldsymbol{\sigma}^1 \coloneqq \operatorname{sign}(\boldsymbol{x}^1), \quad \dots, \quad \boldsymbol{\sigma}^K \coloneqq \operatorname{sign}(\boldsymbol{x}^K).$$

Finally, let $\delta^{K-1} > b_{m+1} > \delta^K$, $0 < 2\alpha\|\boldsymbol{x}^K\|_1 < 1$ and $S_\delta = S_K$ for $\delta^{K-1} > \delta \geq \delta^K$,

$$\tilde{\boldsymbol{A}} \coloneqq \begin{bmatrix} \boldsymbol{A} & 2\alpha\boldsymbol{b} \\ 0 & \alpha b_{m+1} \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{b}} \coloneqq \begin{pmatrix} \boldsymbol{b} \\ b_{m+1} \end{pmatrix}.$$

Then, the problem

$$\min_{(\boldsymbol{x}, x_{n+1}) \in \mathbb{R}^n \times \mathbb{R}} \left\| \begin{pmatrix} \boldsymbol{x} \\ x_{n+1} \end{pmatrix} \right\|_1 \quad \text{s.t.} \quad \left\| \tilde{\boldsymbol{A}} \begin{pmatrix} \boldsymbol{x} \\ x_{n+1} \end{pmatrix} - \tilde{\boldsymbol{b}} \right\|_\infty \leq \eta. \tag{$\tilde{\mathrm{P}}_\eta$}$$

has a unique solution for each $\eta \geq 0$. Moreover, ℓ_1-Houdini applied to the problem ($\tilde{\mathrm{P}}_0$) produces $3(K+1)-1$ primal iterates and the corresponding sign patterns are

$$\underbrace{\begin{pmatrix} \boldsymbol{\sigma}^0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} \boldsymbol{\sigma}^K \\ 0 \end{pmatrix}}_{\text{first } K+1 \text{ patterns}}, \underbrace{\begin{pmatrix} \boldsymbol{\sigma}^{K-1} \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} \boldsymbol{\sigma}^0 \\ 1 \end{pmatrix}}_{\text{middle } K \text{ patterns}}, \underbrace{\begin{pmatrix} -\boldsymbol{\sigma}^0 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} -\boldsymbol{\sigma}^K \\ 1 \end{pmatrix}}_{\text{last } K+1 \text{ patterns}}.$$

Proof. First, we observe that

$$\tilde{\boldsymbol{A}} \begin{pmatrix} \boldsymbol{x} \\ x_{n+1} \end{pmatrix} - \tilde{\boldsymbol{b}} = \begin{pmatrix} \boldsymbol{A}\boldsymbol{x} - (1 - 2\alpha x_{n+1})\boldsymbol{b} \\ -(1 - \alpha x_{n+1}) b_{m+1} \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{A}}^\top \begin{pmatrix} \boldsymbol{y} \\ y_{m+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^\top \boldsymbol{y} \\ 2\alpha \boldsymbol{b}^\top \boldsymbol{y} + \alpha b_{m+1} y_{m+1} \end{pmatrix}.$$

Therewith, it follows from (3.1) that $\left(\begin{pmatrix} \tilde{\boldsymbol{x}} \\ \tilde{x}_{n+1} \end{pmatrix}, \begin{pmatrix} \tilde{\boldsymbol{y}} \\ \tilde{y}_{m+1} \end{pmatrix} \right)$ is an optimal pair for ($\tilde{\mathrm{P}}_\eta$) if and only if the conditions

$$-\boldsymbol{A}^\top \tilde{\boldsymbol{y}} \in \operatorname{Sign}(\tilde{\boldsymbol{x}}) \tag{3.12a}$$

$$-2\alpha \boldsymbol{b}^\top \tilde{\boldsymbol{y}} - \alpha b_{m+1} \tilde{y}_{m+1} \in \operatorname{Sign}(\tilde{x}_{n+1}) \tag{3.12b}$$

and

$$\boldsymbol{A}\tilde{\boldsymbol{x}} - (1 - 2\alpha \tilde{x}_{n+1})\boldsymbol{b} \in \eta \operatorname{Sign}(\tilde{\boldsymbol{y}}) \tag{3.13a}$$

$$-(1 - \alpha \tilde{x}_{n+1}) b_{m+1} \in \eta \operatorname{Sign}(\tilde{y}_{m+1}) \tag{3.13b}$$

are satisfied.

If $\tilde{x}_{n+1} \neq \frac{1}{2\alpha}$, then by dividing (3.12a) and (3.13a) by $\operatorname{sign}(1-2\alpha\tilde{x}_{n+1})$ and $1-2\alpha\tilde{x}_{n+1}$, respectively, we see that

$$\left(\frac{\tilde{\boldsymbol{x}}}{1-2\alpha\tilde{x}_{n+1}}, \frac{\tilde{\boldsymbol{y}}}{\operatorname{sign}(1-2\alpha\tilde{x}_{n+1})}\right) \quad \text{is an optimal pair for} \quad (\mathrm{P}_{\frac{\eta}{|1-2\alpha\tilde{x}_{n+1}|}}). \tag{3.14}$$

If $\tilde{x}_{n+1} = \frac{1}{2\alpha}$, then it follows from (3.12a) and (3.13a) that $\tilde{\boldsymbol{x}} = \mathbf{0}$ and $\tilde{\boldsymbol{y}} = \mathbf{0}$ form an optimal pair for (P_η).

We take up (3.14) again later. For the moment, we keep the following in mind: As either $\tilde{\boldsymbol{y}}$ or $-\tilde{\boldsymbol{y}}$ is a valid dual certificate for (P_δ) for some value of δ, Corollary 32 implies that $|\boldsymbol{b}^\top\tilde{\boldsymbol{y}}| \leq |\boldsymbol{b}^\top\boldsymbol{y}^K|$. Moreover, as the optimal objective function values of (P_0) and (D_0) are equal, we conclude that $|\boldsymbol{b}^\top\boldsymbol{y}^K| = \|\boldsymbol{x}^K\|_1$. Therewith, it follows from our initial assumption on α that $2\alpha|\boldsymbol{b}^\top\tilde{\boldsymbol{y}}| < 1$ regardless of η.

Next, we show that $\tilde{x}_{n+1}$ is uniquely determined for each $\eta \geq 0$. We start with the case $\eta < b_{m+1}$ and use (3.13b) to see that

$$\tilde{x}_{n+1} \in \frac{1}{\alpha}\left(1 + \frac{\eta}{b_{m+1}}\operatorname{Sign}(\tilde{y}_{m+1})\right) \subseteq \left(0, \frac{2}{\alpha}\right)$$

and hence, $\operatorname{sign}(\tilde{x}_{n+1}) = 1$. Rearranging (3.12b) now yields

$$\tilde{y}_{m+1} = \frac{-1 - 2\alpha\boldsymbol{b}^\top\tilde{\boldsymbol{y}}}{\alpha b_{m+1}} < 0.$$

Thus, we have $\operatorname{sign}(\tilde{y}_{m+1}) = -1$ and

$$\tilde{x}_{n+1} = \frac{1}{\alpha}\left(1 - \frac{\eta}{b_{m+1}}\right) \in \left(0, \frac{1}{\alpha}\right].$$

Now, let $\eta \geq b_{m+1}$ and suppose that $\tilde{x}_{n+1} \neq 0$. Analogous to above, we obtain from (3.12b) that $\operatorname{sign}(\tilde{y}_{m+1}) = -\operatorname{sign}(\tilde{x}_{n+1})$ and therewith from (3.13b) that

$$\tilde{x}_{n+1} = \frac{1}{\alpha}\left(1 - \frac{\eta}{b_{m+1}}\operatorname{sign}(\tilde{x}_{n+1})\right).$$

Multiplying both sides with $\operatorname{sign}(\tilde{x}_{n+1})$ gives us

$$|\tilde{x}_{n+1}| = \frac{1}{\alpha}\left(\operatorname{sign}(\tilde{x}_{n+1}) - \frac{\eta}{b_{m+1}}\right) \leq 0$$

which contradicts our assumption that $\tilde{x}_{n+1} \neq 0$. It follows that $\tilde{x}_{n+1}$ is uniquely determined for each $\eta \geq 0$. Therefore, we use the notation $\tilde{x}_{n+1}(\eta)$ and conclude that

$$\tilde{x}_{n+1}(\eta) = \begin{cases} 0 & , \ \eta \geq b_{m+1} \\ \frac{b_{m+1}-\eta}{\alpha b_{m+1}} & , \ \eta < b_{m+1} \end{cases}, \tag{3.15}$$

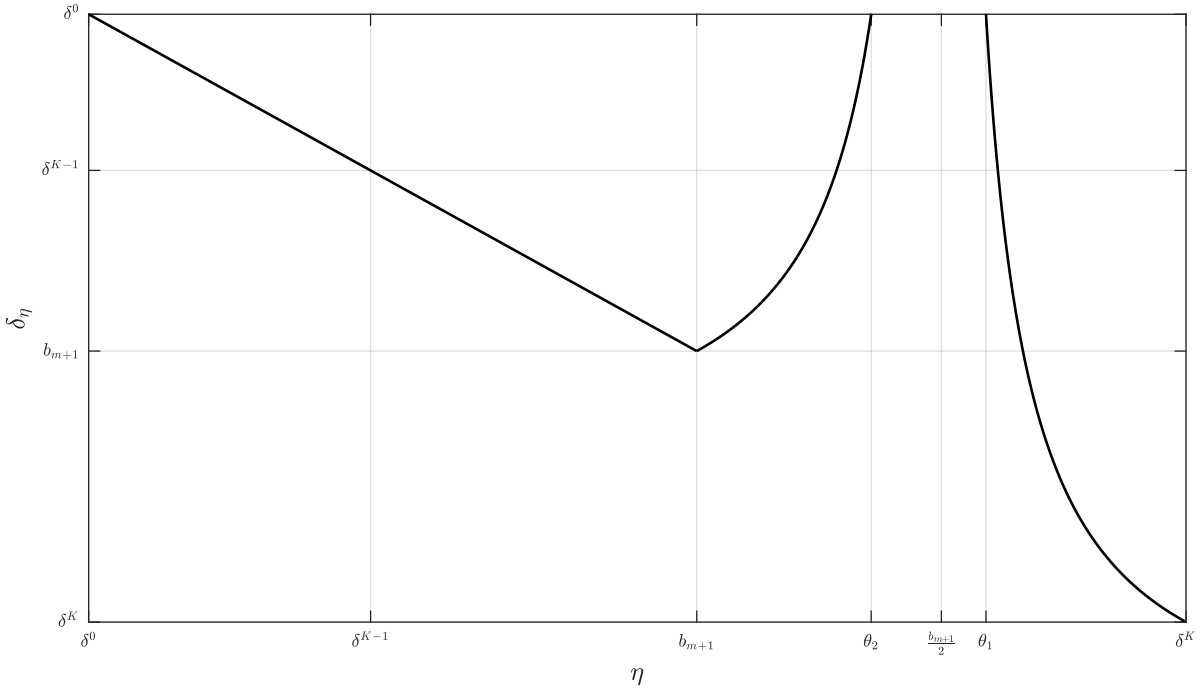

Figure 3.3: The solution path of $(\tilde{\mathrm{P}}_\eta)$ inherits its structure from the solution path of (P_δ) in the sense that $\tilde{\boldsymbol{x}}(\eta) = \pm\frac{\eta}{\delta_\eta}\boldsymbol{x}^\star(\delta_\eta)$. The above plot exemplarily shows the behavior of δ_η according to (3.17). In particular, it is apparent that δ_η traverses the interval $[b_{m+1}, \delta^0]$ three times.

$$1 - 2\alpha\tilde{x}_{n+1}(\eta) = \begin{cases} 1 & , \ \eta \geq b_{m+1} \\ \frac{2\eta - b_{m+1}}{b_{m+1}} & , \ \eta < b_{m+1} \end{cases}, \tag{3.16}$$

and, in case $\eta \neq \frac{b_{m+1}}{2}$,

$$\delta_\eta \coloneqq \frac{\eta}{|1 - 2\alpha\tilde{x}_{n+1}(\eta)|} = \begin{cases} \eta & , \ \eta \geq b_{m+1} \\ \frac{\eta b_{m+1}}{2\eta - b_{m+1}} & , \ b_{m+1} > \eta > \frac{b_{m+1}}{2} \\ \frac{\eta b_{m+1}}{b_{m+1} - 2\eta} & , \ \frac{b_{m+1}}{2} > \eta \geq 0 \end{cases}. \tag{3.17}$$

Moreover, it turns out that

$$\delta_\eta \geq \delta^0 \ \Leftrightarrow \ \eta \in \left[\underbrace{\frac{b_{m+1}\delta^0}{2\delta^0 + b_{m+1}}}_{=:\theta_1}, \underbrace{\frac{b_{m+1}\delta^0}{2\delta^0 - b_{m+1}}}_{=:\theta_2} \right] \setminus \left\{ \frac{b_{m+1}}{2} \right\} \tag{3.18}$$

for $\eta < b_{m+1}$ (see Figure 3.3).

As $\tilde{x}_{n+1}(\eta)$ is uniquely determined and $\boldsymbol{x}^\star(\delta)$ is the unique solution of (P_δ) for each

$\delta \geq 0$, it follows from (3.14)–(3.18) that

$$\tilde{\boldsymbol{x}}(\eta) = \begin{cases} \boldsymbol{x}^\star(\eta) & , \ \eta \geq b_{m+1} \\ \frac{2\eta - b_{m+1}}{b_{m+1}} \boldsymbol{x}^\star\left(\frac{\eta b_{m+1}}{2\eta - b_{m+1}}\right) & , \ b_{m+1} > \eta > \theta_2 \\ \mathbf{0} & , \ \theta_2 \geq \eta \geq \theta_1 \\ \frac{2\eta - b_{m+1}}{b_{m+1}} \boldsymbol{x}^\star\left(\frac{\eta b_{m+1}}{b_{m+1} - 2\eta}\right) & , \ \theta_1 > \eta \geq 0 \end{cases} \tag{3.19}$$

and $\tilde{x}_{n+1}(\eta)$ form the unique solution of $(\tilde{\mathrm{P}}_\eta)$ for each $\eta \geq 0$.

By Lemma 24, the solution path of $(\tilde{\mathrm{P}}_\eta)$ is continuous piecewise linear. Moreover, (3.19) makes clear that the solution path of $(\tilde{\mathrm{P}}_\eta)$, up to the last component $\tilde{x}_{n+1}(\eta)$, inherits its structure from the solution path of (P_δ). As a consequence of Lemma 17, the iterates produced by ℓ_1-HOUDINI correspond exactly to the kinks in the solution path of $(\tilde{\mathrm{P}}_\eta)$. In the following, we specify the locations of the kinks and the associated sign patterns of $\tilde{\boldsymbol{x}}(\eta)$ and $\tilde{x}_{n+1}(\eta)$.

Let us start with the case $\eta \geq b_{m+1}$. By (3.19), we have $\tilde{\boldsymbol{x}}(\eta) = \boldsymbol{x}^\star(\eta)$ and $\tilde{x}_{n+1}(\eta) = 0$. Since $\delta^{K-1} > b_{m+1} > \delta^K$, it follows that ℓ_1-HOUDINI produces the iterates $(\boldsymbol{x}^k, 0)$ as well as the related sign patterns $(\boldsymbol{\sigma}^k, 0)$ for $k = 0, \ldots, K-1$. The fact that $\tilde{x}_{n+1}(\eta) = 0$ for $\eta \geq b_{m+1}$ and $\tilde{x}_{n+1}(\eta) > 0$ for $\eta < b_{m+1}$ implies that the next kink is located at $(\boldsymbol{x}^\star(b_{m+1}), 0)$ and because $S_\delta = S_0$ for $\delta < \delta^{K-1}$, the associated sign pattern is $(\boldsymbol{\sigma}^K, 0)$.

Second, we consider the case $b_{m+1} > \eta > \theta_2$. With respect to (3.19), it holds that $(2\eta - b_{m+1})/b_{m+1} > 0$. Moreover, the term $\eta b_{m+1}/(2\eta - b_{m+1})$ is strictly monotonically increasing in η and takes on each value in the interval (b_{m+1}, δ^0). As $\eta b_{m+1}/(2\eta - b_{m+1}) = \delta^k$ if and only if $\eta = \delta^k b_{m+1}/(2\delta^k - b_{m+1})$, it follows that the next $K-1$ kinks are located at

$$\left(\frac{b_{m+1}}{2\delta^k - b_{m+1}} \boldsymbol{x}^\star(\delta^k), \frac{\delta^k - b_{m+1}}{\alpha(2\delta^k - b_{m+1})}\right) \tag{3.20}$$

with related sign patterns $(\boldsymbol{\sigma}^k, 1)$ for $k = K-1, \ldots, 1$.

Thirdly, for $\theta_2 \geq \eta \geq \theta_1$, we obtain that the next two kinks are located at

$$\left(\mathbf{0}, \frac{b_{m+1} - \theta_2}{\alpha b_{m+1}}\right) \quad \text{and} \quad \left(\mathbf{0}, \frac{b_{m+1} - \theta_1}{\alpha b_{m+1}}\right) \tag{3.21}$$

because these two points bound one linear segment of the solution path. The associated sign patterns are both $(\mathbf{0}, 1) = (\boldsymbol{\sigma}^0, 1)$.

The remaining case ist $\theta_1 > \eta \geq 0$ which is analogous to the second case. Here, it holds that $(2\eta - b_{m+1})/b_{m+1} < 0$. Further, the term $\eta b_{m+1}/(b_{m+1} - 2\eta)$ is strictly monotonically decreasing in η and takes on each value in the range $(\delta^0, 0]$. Since $\eta b_{m+1}/(b_{m+1} - 2\eta) = \delta^k$ if and only if

$$\eta = \frac{b_{m+1}\delta^k}{2\delta^k + b_{m+1}} \ , \tag{3.22}$$

we obtain that the next K kinks are located at

$$\left(-\frac{b_{m+1}}{2\delta^k + b_{m+1}}\boldsymbol{x}^\star(\delta^k), \frac{\delta^k + b_{m+1}}{\alpha(2\delta^k + b_{m+1})}\right) \tag{3.23}$$

and that the related sign patterns are $(-\boldsymbol{\sigma}^k, 1)$ for $k = 1, \ldots, K$. □

Remark 36. We recall that $S_\delta = S_K$ for $\delta \in [\delta^K, \delta^{K-1})$ if and only if the support of $\boldsymbol{x}^\star(\delta)$ is equal to the support of $\boldsymbol{x}^K = \boldsymbol{x}^\star(0)$ for the respective values of the homotopy parameter, i.e., the optimal support does not change on the last linear segment of the solution path (excluding the second-to-last kink $\boldsymbol{x}^{K-1}$). Actually, dropping this assumption from Proposition 35 would not change the statement much. In Lemma 28, we have already seen that the support S_δ is invariant on (δ^K, δ^{K-1}), whereas in general, it is at least possible that $S_K \subset S_\delta$. This would be the case if and only if, simultaneously with the last inactive constraint(s) becoming active, one or more non-zero components of $\boldsymbol{x}^\star(\delta)$ become zero as soon as δ is driven to zero. Either way, the number of iterations needed to solve $(\tilde{\mathrm{P}}_\eta)$ (and with that, the number of kinks in the associated solution path) is not affected. Only, in the sign pattern $\boldsymbol{\sigma}^K$ that is claimed to appear twice in the solution path of $(\tilde{\mathrm{P}}_\eta)$, we would have to replace one or more non-zero components by zeros.

Remark 37. In the proof of Proposition 35, we have seen that the solution path of $(\tilde{\mathrm{P}}_\eta)$ inherits most of its kinks from the solution path of (P_δ). More precisely, each single kink associated to one of the values $\delta^1, \ldots, \delta^{K-1}$ induces three new kinks in the solution path of $(\tilde{\mathrm{P}}_\eta)$, which amounts to $3(K-1)$ kinks. The exact locations of these new kinks can be calculated explicitly by inverting all three cases in (3.17) separately and plugging in $\delta^1, \ldots, \delta^{K-1}$ afterwards. In addition, there are five more new kinks which are directly associated to $\delta^0, b_{m+1}, \theta_2, \theta_1$ and δ^K (in this order).

Theorem 38. *In the worst case,* ℓ_1-HOUDINI *produces at least* $(3^n + 1)/2$ *iterates.*

Proof. From Proposition 35, we deduce the following *resisting strategy*: Starting with $n = 1$ and setting $\boldsymbol{A} = 1$ as well as $\boldsymbol{b} = 1$, we see that the associated problem (P_δ) has a unique solution for each $\delta \geq 0$. In particular, the solution path of (P_δ) has $2 = (3^1 + 1)/2$ linear segments and ℓ_1-HOUDINI produces the iterates $\boldsymbol{x}^0 = 0$ and $\boldsymbol{x}^1 = 1$ (corresponding to the kinks of the solution path) and the related values $\delta^0 = 1$ and $\delta^1 = 0$ of the homotopy parameter. Building on this first instance of (P_δ) and consecutively constructing $(\tilde{\mathrm{P}}_\eta)$ according to Proposition 35, we arrive at an instance of (P_δ) with $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$ after $n - 1$ steps. Now, suppose that $n > 1$ and that the statement holds true for dimension $n - 1$. Then, it follows from Proposition 35 that ℓ_1-HOUDINI applied to the n-dimensional instance generates $3[(3^{n-1}+1)/2] - 1 = (3^n + 1)/2$ iterates. □

Remark 39. The term *at least* in Theorem 38 refers to the circumstance that it is unknown whether the described resisting strategy is actually a worst-case example. This is due to the fact that the number of $(3^n + 1)/2$ iterates is smaller than the upper bound of $(3^{m+n} + 1)/2$ iterations which we established in Theorem 34. Regarding that the constructed example does still induce a computational complexity that is exponential

in the number of variables, it is probably more appropriate to use the term *bad-case example*.

In the following, we further investigate the specific type of instances that emerge if we repeatedly apply the transition from (P_δ) to $(\tilde{\mathrm{P}}_\eta)$. In that context $(\mathrm{P}_\delta)^{(n)}$ denotes the optimization problem associated to $\boldsymbol{A}^{(n)} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b}^{(n)} \in \mathbb{R}^n$. Moreover, $\boldsymbol{x}^{(n)}$ denotes the (unique) optimal solution of the problem $(\mathrm{P}_0)^{(n)}$.

Proposition 40. *Let $\boldsymbol{A}^{(1)} = \alpha_1 \in \mathbb{R}_+$ and $\boldsymbol{b}^{(1)} = b_1 \in \mathbb{R}_+$. For $n \geq 2$, let further $\boldsymbol{A}^{(n)} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b}^{(n)} \in \mathbb{R}^n$ be constructed recursively according to the strategy described in the proof of Theorem 38, i.e.,*

$$\boldsymbol{A}^{(n)} \coloneqq \begin{bmatrix} \boldsymbol{A}^{(n-1)} & 2\alpha_n \boldsymbol{b}^{(n-1)} \\ 0 & \alpha_n b_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}^{(n)} \coloneqq \begin{pmatrix} \boldsymbol{b}^{(n-1)} \\ b_n \end{pmatrix}, \tag{3.24}$$

where $0 < b_n < \lambda^{(n-1)}$ is satisfied for the smallest non-zero homotopy parameter $\lambda^{(n-1)}$ associated to a kink on the solution path of $(\mathrm{P}_\delta)^{(n-1)}$, and $0 < 2\alpha_n \|\boldsymbol{x}^{(n-1)}\|_1 < 1$ is valid for the optimal solution $\boldsymbol{x}^{(n-1)}$ of $(\mathrm{P}_0)^{(n-1)}$. Then, it holds that

$$\boldsymbol{A}^{(n)} = \begin{bmatrix} \alpha_1 & 2\alpha_2 b_1 & 2\alpha_3 b_1 & \cdots & \cdots & 2\alpha_n b_1 \\ 0 & \alpha_2 b_2 & 2\alpha_3 b_2 & \cdots & \cdots & 2\alpha_n b_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \alpha_{n-1} b_{n-1} & 2\alpha_n b_{n-1} \\ 0 & \cdots & \cdots & \cdots & 0 & \alpha_n b_n \end{bmatrix}. \tag{3.25}$$

Further, $\boldsymbol{x}^{(1)} = \frac{b_1}{\alpha_1}$ and there exist $0 < p_2, \ldots, p_n < 1$ such that, with $\gamma \in \mathbb{R}^{n-1}$ defined recursively as

$$\gamma_1 \coloneqq 1 \quad \text{and} \quad \gamma_k \coloneqq \frac{2 + p_k}{p_k} \gamma_{k-1} \quad \text{for} \quad k = 2, \ldots, n\ , \tag{3.26}$$

it holds that

$$\alpha_n = \frac{p_n}{2\gamma_{n-1}\frac{b_1}{\alpha_1}}\ , \quad \boldsymbol{x}^{(n)} = \begin{pmatrix} -\boldsymbol{x}^{(n-1)} \\ \frac{1}{\alpha_n} \end{pmatrix} \quad \text{and} \quad \left\|\boldsymbol{x}^{(n)}\right\|_1 = \gamma_n \frac{b_1}{\alpha_1}. \tag{3.27}$$

Finally, there exist $0 < q_2, \ldots, q_n < 1$ such that

$$\lambda^{(n-1)} = \frac{b_1}{\prod_{k=2}^{n-1}\left(1 + \frac{2}{q_k}\right)} \quad \text{and} \quad b_n = q_n \lambda^{(n-1)}. \tag{3.28}$$

Proof. By applying (3.24) recursively, it follows immediately that the matrix $\boldsymbol{A}^{(n)}$ has the claimed structure (3.25).

In case $n = 1$, the associated optimal solution is given by

$$\boldsymbol{x}^{(1)} = \operatorname*{arg\,min}_{x\in\mathbb{R}} \ |x| \quad \text{s.t. } \alpha_1 x - b_1 = 0.$$

It follows that $\boldsymbol{x}^{(1)} = \frac{b_1}{\alpha_1} = \gamma_1 \frac{b_1}{\alpha_1}$. By assumption, we further have $0 < 2\alpha_2 \|\boldsymbol{x}^{(1)}\|_1 < 1$ and thus, there exists a $p_2 \in (0,1)$ such that $\alpha_2 = p_2/(2\gamma_1 \frac{b_1}{\alpha_1})$.

Now, suppose that $\|\boldsymbol{x}^{(n-1)}\|_1 = \gamma_{n-1} \frac{b_1}{\alpha_1}$ and $\alpha_n = p_n/(2\gamma_{n-1} \frac{b_1}{\alpha_1})$ for some fixed $n \geq 2$. From (3.15) and (3.19) (with $\eta = 0$), it follows that $\boldsymbol{x}^{(n)} = (-\boldsymbol{x}^{(n-1)}, \frac{1}{\alpha_n})$ and hence,

$$\|\boldsymbol{x}^{(n)}\|_1 = \gamma_{n-1} \frac{b_1}{\alpha_1} + \frac{1}{\alpha_n} = \gamma_{n-1} \frac{b_1}{\alpha_1} + \frac{2\gamma_{n-1} \frac{b_1}{\alpha_1}}{p_n} = \frac{2 + p_n}{p_n} \gamma_{n-1} \frac{b_1}{\alpha_1} = \gamma_n \frac{b_1}{\alpha_1}.$$

Moreover, we need to choose α_{n+1} according to $0 < 2\alpha_n \|\boldsymbol{x}^{(n)}\|_1 < 1$ and conclude that there exists a $p_{n+1} \in (0,1)$ such that

$$\alpha_{n+1} = \frac{p_{n+1}}{2\|\boldsymbol{x}^{(n)}\|_1} = \frac{p_{n+1}}{2\gamma_n \frac{b_1}{\alpha_1}}.$$

Up to this point, we have shown that the first set of statements, from (3.25)–(3.27), is true. To see that the last statement is true as well, we start with the observation that the solution path of $(\mathrm{P}_\delta)^{(0)}$ has two linear segments that meet at one kink corresponding to the value $\|\boldsymbol{b}^{(1)}\|_\infty = b_1$ of the homotopy paramter. As a consequence, it holds that $\lambda^{(1)} = b_1$. Since we further require $0 < b_2 < \lambda^{(1)}$, we conclude that there exists a $q_2 \in (0,1)$ such that $b_2 = q_2 \lambda^{(1)} = q_2 b_1$. Using (3.22), we conclude that

$$\lambda^{(2)} = \frac{b_2 \lambda^{(1)}}{2\lambda^{(1)} + b_2} = \frac{q_2 b_1 \lambda^{(1)}}{2\lambda^{(1)} + q_2 \lambda^{(1)}} = \frac{b_1}{1 + \frac{2}{q_2}}.$$

Finally, suppose that

$$\lambda^{(n-1)} = \frac{b_1}{\prod_{k=2}^{n-1} \left(1 + \frac{2}{q_k}\right)}$$

holds for some fixed $n \geq 3$. Analogous to above, there exists a $q_n \in (0,1)$ such that $b_n = q_n \lambda^{(n-1)}$. Using (3.22) again, we obtain that

$$\lambda^{(n)} = \frac{b_n \lambda^{(n-1)}}{2\lambda^{(n-1)} + b_n} = \frac{\lambda^{(n-1)}}{2\frac{\lambda^{(n-1)}}{b_n} + 1} = \frac{\lambda^{(n-1)}}{\frac{2}{q_n} + 1} = \frac{b_1}{\prod_{k=2}^{n} \left(1 + \frac{2}{q_k}\right)}.$$

□

In particular, Proposition 40 reveals that each possible combination of $\boldsymbol{A}^{(n)}$ and $\boldsymbol{b}^{(n)}$ can be constructed without explicitly taking care of the conditions $0 < b_k < \lambda^{(k-1)}$ and $0 < 2\alpha_k \|\boldsymbol{x}^{(k-1)}\| < 1$ before each stage of construction. To the contrary, we only need to choose $\alpha_1, b_1 \in \mathbb{R}_+$ as well as $p_2, \ldots, p_n, q_2, \ldots, q_n \in (0,1)$ in advance. Then, successively constructing $\boldsymbol{A}^{(k)}$ and $\boldsymbol{b}^{(k)}$ ($n-1$ times) according to (3.24)–(3.28) finally yields the associated instance $\boldsymbol{A}^{(n)}$ and $\boldsymbol{b}^{(n)}$.

Although it is possible by Proposition 40 to construct $\boldsymbol{A}^{(n)}$ and $\boldsymbol{b}^{(n)}$ for arbitrarily large $n \in \mathbb{N}$, the following Corollary 41 shows that the absolute value of the smallest

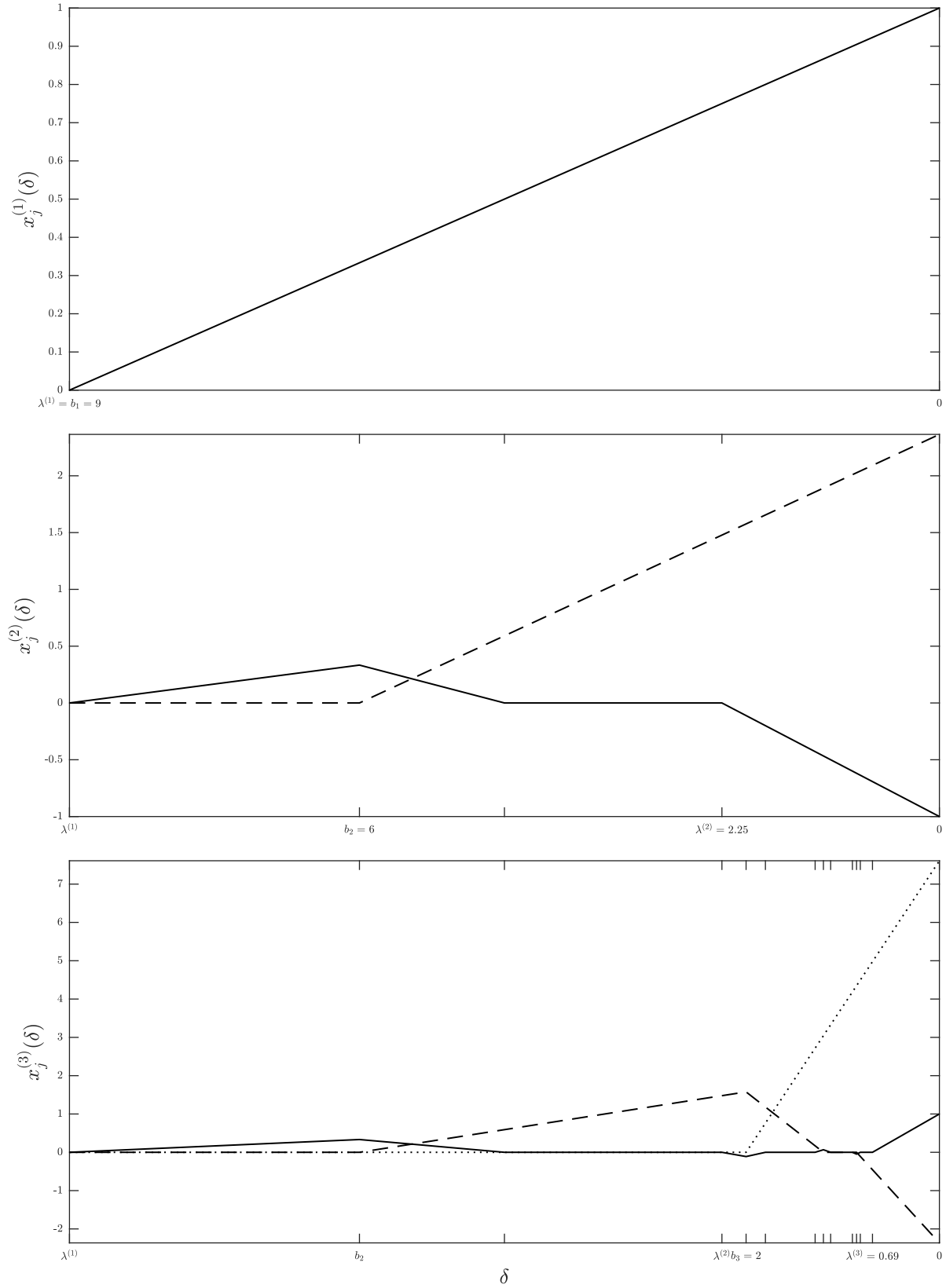

Figure 3.4: Solution paths generated by ℓ_1-HOUDINI applied to the problems $(\mathrm{P}_0)^{(k)}$ for $k = 1, 2, 3$, where $\boldsymbol{A}^{(k)}$ and $\boldsymbol{b}^{(k)}$ were constructed according to Proposition 40 with $\alpha_1 = b_1 = 3^{n-1}$, $q_2 = 2/3$, $q_3 = 8/9$ and p_1, p_2 drawn randomly from the interval (0.5, 1).

entries in $\boldsymbol{A}^{(n)}$ and $\boldsymbol{b}^{(n)}$ decreases exponentially fast in the number of variables. At the same time, the optimal objective function value of $(\mathrm{P}_0)^{(n)}$, i.e., the ℓ_1-norm of the associated solution $\boldsymbol{x}^{(n)}$, grows exponentially fast.

Corollary 41. *Under the same prerequisites as in Proposition 40, it holds that*

$$\alpha_n < 3^{-(n-2)} \cdot \frac{1}{2\frac{b_1}{\alpha_1}} \ , \quad x_n^{(n)} > 3^{n-2} \cdot 2\frac{b_1}{\alpha_1} \ , \quad \|\boldsymbol{x}^{(n)}\|_1 > 3^{n-1} \cdot \frac{b_1}{\alpha_1} \tag{3.29}$$

and

$$b_n < \lambda^{(n-1)} < 3^{-(n-2)} \cdot b_1. \tag{3.30}$$

Proof. Using $\gamma_1 = 1$ and $0 < p_2, \dots, p_k < 1$, we conclude that

$$\gamma_k = \frac{2 + p_k}{p_k}\gamma_{k-1} > 3\gamma_{k-1} > \cdots > 3^{k-1}\gamma_1 = 3^{k-1}$$

for $k = 2, \dots, n$. Therewith, (3.29) and (3.30) follow immediately from (3.27) and (3.28), respectively. □

Remark 42. As a consequence of Corollary 41, if α_1 and b_1 are fixed, the vast majority of entries in $\boldsymbol{A}^{(n)}$ and $\boldsymbol{b}^{(n)}$ falls below machine precision if we keep increasing n. The situation is slightly different in case n is fixed in advance. Say, we would like to obtain $b_n \approx 1$. To that end, we could simply start with $b_1 = 3^{n-2}$ and set $q_2, \dots, q_n \approx 1$. In turn, we have $\alpha_n < 3^{-2(n-2)} \cdot \alpha_1/2$ by (3.29). In order to obtain $\boldsymbol{A}_{n,n}^{(n)} = \alpha_n b_n \approx 1$ as well, we could set $p_2, \dots, p_n \approx 1$ and choose α_1 such that

$$3^{-2(n-2)} \cdot \frac{\alpha_1}{2} = 1 \Leftrightarrow \alpha_1 = 2 \cdot 3^{2(n-2)}.$$

This technique of course implies that $\boldsymbol{A}_{1,1}^{(n)} = \alpha_1$ is large when n is large. Moreover, it follows that

$$x_1^{(n)} = \frac{b_1}{\alpha_1} = \frac{3^{n-2}}{2 \cdot 3^{2(n-2)}} = \frac{1}{2} \cdot 3^{-(n-2)}$$

and hence, the leading entries of $\boldsymbol{x}^{(n)}$ fall below machine precision as soon as n is sufficiently large. At the same time, it holds that

$$x_n^{(n)} \approx 3^{n-2} \cdot 2\frac{b_1}{\alpha_1} \approx 1 \quad \text{and} \quad \|\boldsymbol{x}^{(n)}\|_1 \approx 3^{n-1} \cdot \frac{b_1}{\alpha_1} = \frac{3}{2}.$$

Example 43. We construct a five-dimensional example proceeding as described in Remark 42 with $\alpha_1 = 2 \cdot 3^6$, $b_1 = 3^3$ and $p_k = q_k = 1 - 10^{-6}$ for $k = 2, \dots, 5$. Therewith, we obtain $\lambda^{(k)} \approx 3^{4-k}$, $\gamma_k \approx 3^{k-1}$ and $\alpha_k \approx 3^{5-k}$ for $k = 2, \dots, 5$.

$$\boldsymbol{A}^{(5)} \approx \begin{bmatrix} 2 \cdot 3^6 & 2 \cdot 3^6 & 2 \cdot 3^5 & 2 \cdot 3^4 & 2 \cdot 3^3 \\ 0 & 3^6 & 2 \cdot 3^5 & 2 \cdot 3^4 & 2 \cdot 3^3 \\ 0 & 0 & 3^4 & 2 \cdot 3^3 & 2 \cdot 3^2 \\ 0 & 0 & 0 & 3^2 & 2 \cdot 3^1 \\ 0 & 0 & 0 & 0 & 3^0 \end{bmatrix} \ , \quad \boldsymbol{b}^{(5)} \approx \begin{pmatrix} 3^3 \\ 3^3 \\ 3^2 \\ 3^1 \\ 3^0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{x}^{(5)} \approx \begin{pmatrix} \frac{1}{2} \cdot 3^{-3} \\ -3^{-3} \\ 3^{-2} \\ -3^{-1} \\ 3^0 \end{pmatrix}.$$

The numbers on the respective right-hand sides correspond to the case $p_k = q_k = 1$ for $k = 1, \ldots, 5$. Note that, although the ℓ_2-distance to the true values of $\boldsymbol{A}^{(5)}$, $\boldsymbol{b}^{(5)}$ and $\boldsymbol{x}^{(5)}$ is relatively small (less than 10^{-2} in each case), these approximate values do not depict an example in the sense of Proposition 40. Indeed, using the approximate values, we obtain a solution path with 42 linear segments, whereas Theorem 38 predicts 122 segments for an appropriate five-dimensional example.

4 Active-Set Methods

Each iteration of our ℓ_1-HOUDINI algorithm (see Algorithm 1) requires the solutions of two linear programs, one for the dual update and one for the primal update. In principle, an arbitrary LP solver can be used to tackle these linear programs. This can be considered a feature of our method which makes it clear and easy to implement. However, there is a specific structure underlying the dual and primal update problems (U_D^k) and (U_P^k), respectively. First, the previous dual iterate $\boldsymbol{y}^k$ is always feasible for (U_D^k) and the previous primal iterate $\boldsymbol{x}^k$ is always feasible for (U_P^k). Second, we have seen in Examples 16 and 18 that both updates can, at least a posteriori, be expressed in terms of improvement directions: In view of the dual update, there exists a direction $\boldsymbol{e}^{k+1}$ such that $(\boldsymbol{x}^k, \boldsymbol{y}^k + s\boldsymbol{e}^{k+1})$ is an optimal pair for (P_{δ^k}) for $0 \leq s \leq 1$ and the optimum in (U_D^k) is attained at $s = 1$. Analogously, with respect to the primal update, we have seen that there exists a direction $\boldsymbol{d}^{k+1}$ such that $(\boldsymbol{x}^k + t\boldsymbol{d}^{k+1}, \boldsymbol{y}^{k+1})$ is an optimal pair for $(\mathrm{P}_{\delta^k - t})$ for $0 \leq t \leq t^{k+1}$ and the optimum in (U_P^k) is obtained with $t = t^{k+1}$. Using the example of the dual update, these two characteristics raise the question whether, instead of solving (U_D^k) from scratch in each iteration, it would also work to start at the previous iterate $\boldsymbol{y}^k$ and find an appropriate direction $\boldsymbol{e}^{k+1}$ in which to go as far as possible to find an optimal solution $\boldsymbol{y}^{k+1}$. Our hope is, of course, that such an approach would significantly reduce the required computational effort.

In view of Lemma 15, there are two possible approaches in order to identify a direction $\boldsymbol{e}^{k+1}$ and an associated step size s^{k+1}. The first one is to substitute $\boldsymbol{y} =: \boldsymbol{y}^k + s\boldsymbol{e}$ in (U_D^k) (which is problem (3.3) with $\hat{\boldsymbol{x}} = \boldsymbol{x}^k$) and solve the resulting optimization problem for optimal values of s and $\boldsymbol{e}$. As the substitution increases the number of variables from $|W_k|$ to $|W_k| + 1$ (all of which must obey sign constraints), leaves the number of $2n - |S_k|$ equations and inequalities unchanged and even makes the problem non-linear, this idea is seemingly not of any advantage. As opposed to this, the second approach is to determine $\boldsymbol{e}^{k+1}$ as a solution of (3.2) with $\hat{\boldsymbol{x}} = \boldsymbol{x}^k$ and afterwards s^{k+1} as the largest step size such that $(\boldsymbol{x}^k, \boldsymbol{y}^k + s^{k+1}\boldsymbol{e}^{k+1})$ is an optimal pair for (P_{δ^k}). At first glance, this approach seems to be beneficial because the system (3.2) has $|W_k|$ variables (whereof $|W_k| - |\Omega_k|$ have to obey sign constraints) but only $|\Sigma_k| + 1$ equations and inequalites. Unfortunately, there is a rub. The direction $\boldsymbol{e}^{k+1}$ is not necessarily the unique solution of the system (3.2). Hence, as we do not search for the direction and the step size s^{k+1} simultaneously, there might exist a different direction $\boldsymbol{e}$ with associated step size s such that $\boldsymbol{\psi}^\top(\boldsymbol{y}^k + s\boldsymbol{e}) < \boldsymbol{\psi}^\top(\boldsymbol{y}^k + s^{k+1}\boldsymbol{e}^{k+1})$, i.e., $\boldsymbol{y}^k + s^{k+1}\boldsymbol{e}^{k+1}$ is not an optimal solution of the dual update problem (U_D^k).

Our final goal in this chapter is to derive a method for (U_D^k) and (U_P^k) that gets along with relatively small subproblems, as in case of the last-mentioned approach with

only $|\Sigma_k| + 1$ linear constraints in case of the dual update problem, and that yields an optimal solution of the respective problems such that the convergence of ℓ_1-HOUDINI is guaranteed. To that end, we derive an active-set method for linear programs in the following section. After that, we apply this active-set method to the problems (U_D^k) and (U_P^k). We will see that, in terms of the dual update, the resulting method is similar to the above-metioned idea to make a step towards a direction according to (3.3). One major difference is that our active-set approach allows for subsequent steps in more than one direction. This avoids the difficulty that the solution of (3.2) may not be unique. Further, it will be sufficient to deal with linear equation systems (instead of a system of linear equalities and inequalities) in order to generate improvement directions.

The results presented in this chapter have already been published in [5], with the involvement of the author.

4.1 Active-Set Method for Linear Programs

4.1.1 Optimality Conditions

Let $\boldsymbol{\gamma} \in \mathbb{R}^d$, $\boldsymbol{\Phi} \in \mathbb{R}^{k \times d}$, $\boldsymbol{u} \in \mathbb{R}^k$, $\boldsymbol{\Psi} \in \mathbb{R}^{l \times d}$, $\boldsymbol{v} \in \mathbb{R}^l$ and $\boldsymbol{\sigma} \in \{\pm 1\}^d$. We consider the linear program

$$\begin{aligned} \min_{\boldsymbol{z} \in \mathbb{R}^d} \quad & \boldsymbol{\gamma}^\top \boldsymbol{z} \\ \text{s.t.} \quad & \boldsymbol{\Phi}\boldsymbol{z} = \boldsymbol{u} \\ & \boldsymbol{\Psi}\boldsymbol{z} \geq \boldsymbol{v} \\ & \mathrm{Diag}(\boldsymbol{\sigma})\boldsymbol{z} \geq \boldsymbol{0} \end{aligned} \tag{4.1}$$

and assume that it is feasible and bounded. By the well-known *KKT conditions* (see, e.g., [35, Theorem 12.1]), $\boldsymbol{z}^\star$ is an optimal solution of (4.1) if and only if there exist *Lagrange multipliers* $\boldsymbol{\lambda} \in \mathbb{R}^k$, $\boldsymbol{\mu} \in \mathbb{R}^l$ and $\boldsymbol{\nu} \in \mathbb{R}^d$ such that the following conditions

hold:[1]

$$\boldsymbol{\Phi}\boldsymbol{z}^\star = \boldsymbol{u} \tag{4.2a}$$
$$\boldsymbol{\Psi}\boldsymbol{z}^\star \geq \boldsymbol{v} \tag{4.2b}$$
$$\mathrm{Diag}(\boldsymbol{\sigma})\boldsymbol{z}^\star \geq \mathbf{0} \tag{4.2c}$$
$$\boldsymbol{\Phi}^\top\boldsymbol{\lambda} + \boldsymbol{\Psi}^\top\boldsymbol{\mu} + \mathrm{Diag}(\boldsymbol{\sigma})\boldsymbol{\nu} = \boldsymbol{\gamma} \tag{4.2d}$$
$$\boldsymbol{\mu} \odot (\boldsymbol{\Psi}\boldsymbol{z}^\star - \boldsymbol{v}) = \mathbf{0} \tag{4.2e}$$
$$\boldsymbol{\nu} \odot \boldsymbol{z}^\star = \mathbf{0} \tag{4.2f}$$
$$\boldsymbol{\mu} \geq \mathbf{0} \tag{4.2g}$$
$$\boldsymbol{\nu} \geq \mathbf{0}. \tag{4.2h}$$

4.1.2 General Theme

Suppose that $\boldsymbol{z}^\ell \in \mathbb{R}^d$ is feasible for (4.1), i.e., it satisfies (4.2a)–(4.2c). Then, there exist subsets $\mathcal{A} \subseteq \{1, \ldots, l\}$ and $\mathcal{S} \subseteq \{1, \ldots, d\}$ such that

$$\boldsymbol{\Psi}^{\mathcal{A}}\boldsymbol{z}^\ell = \boldsymbol{v}^{\mathcal{A}}, \quad \boldsymbol{\Psi}^{\mathcal{A}^c}\boldsymbol{z}^\ell > \boldsymbol{v}^{\mathcal{A}^c}, \quad \boldsymbol{z}^\ell_{\mathcal{S}^c} = \mathbf{0} \quad \text{and} \quad |\boldsymbol{z}^\ell_{\mathcal{S}}| > \mathbf{0}. \tag{4.3}$$

We refer to $\mathcal{A}$ as the *active set* and further to $\mathcal{S}$ as the *support* of $\boldsymbol{z}^\ell$. In the context of (4.2e) and (4.2f), necessarily $\boldsymbol{\mu}_{\mathcal{A}^c} = \mathbf{0}$ and $\boldsymbol{\nu}_{\mathcal{S}} = \mathbf{0}$ in case $\boldsymbol{z}^\ell$ is an optimal solution of (4.1). The following Lemma exploits this fact and provides alternative optimality conditions for (4.1).

Lemma 44. *A point $\boldsymbol{z}^\ell$ is an optimal solution of (4.1) if and only if it is feasible and there exist $\boldsymbol{\lambda} \in \mathbb{R}^k$ and $\boldsymbol{\mu}_{\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$ such that*

$$\boldsymbol{\Phi}_{\mathcal{S}}^\top\boldsymbol{\lambda} + (\boldsymbol{\Psi}_{\mathcal{S}}^{\mathcal{A}})^\top\boldsymbol{\mu}_{\mathcal{A}} = \boldsymbol{\gamma}_{\mathcal{S}} \tag{4.4a}$$
$$\mathrm{Diag}(\boldsymbol{\sigma}_{\mathcal{S}^c})(\boldsymbol{\gamma}_{\mathcal{S}^c} - \boldsymbol{\Phi}_{\mathcal{S}^c}^\top\boldsymbol{\lambda} - (\boldsymbol{\Psi}_{\mathcal{S}^c}^{\mathcal{A}})^\top\boldsymbol{\mu}_{\mathcal{A}}) \geq \mathbf{0} \quad \textit{and} \tag{4.4b}$$
$$\boldsymbol{\mu}_{\mathcal{A}} \geq \mathbf{0}. \tag{4.4c}$$

Proof. It can easily be seen that the conditions (4.4a)–(4.4c) are equivalent to conditions (4.2d)–(4.2h) with $\boldsymbol{\mu}_{\mathcal{A}^c} = \mathbf{0}$, $\boldsymbol{\nu}_{\mathcal{S}} = \mathbf{0}$ and

$$\boldsymbol{\nu}_{\mathcal{S}^c} = \mathrm{Diag}(\boldsymbol{\sigma}_{\mathcal{S}^c})(\boldsymbol{\gamma}_{\mathcal{S}^c} - \boldsymbol{\Phi}_{\mathcal{S}^c}^\top\boldsymbol{\lambda} - (\boldsymbol{\Psi}_{\mathcal{S}^c}^{\mathcal{A}})^\top\boldsymbol{\mu}_{\mathcal{A}}). \tag{4.5}$$

□

[1] The KKT conditions are necessary conditions for a local optimum $\boldsymbol{z}^\star$ in case the objective function as well as the constraints are continuously differentiable. In linear programming, it is rather common to use the term *complementary slackness* for the same type of optimality conditions (see, e.g., [29, Theorems 1 & 2]), and in this case the conditions are actually necessary *and* sufficient for $\boldsymbol{z}^\star$ to be a global optimum. Nevertheless, we use the term *KKT conditions* as it is frequently used in connection with the denotation *Lagrange multipliers* for the associated vectors.

Lemma 44 can be beneficial when we want to check whether a given point $\boldsymbol{z}^\ell$ is an optimal solution of (4.1) or not. In that case, we can first try to find a solution to the system (4.4a) which has $|\mathcal{S}|$ equations, and afterwards verify (4.4b) and (4.4c). In contrast, applying (4.2e)–(4.2h) entails a system with n equations.

Starting from $\boldsymbol{z}^\ell$, our goal is to approach a solution of (4.1) by generating *descent directions* $\boldsymbol{\zeta}$ that preserve the active set as well as the support and, should this not be possible, by changing these sets appropriately. We repeat these steps until we finally identify $\mathcal{A}$, $\mathcal{S}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}_\mathcal{A}$ satisfying (4.4a)–(4.4c).

4.1.3 Descent Directions and Blocking Constraints

If there exists a solution of the linear system

$$\begin{bmatrix} \boldsymbol{\Phi}_\mathcal{S} \\ \boldsymbol{\Psi}_\mathcal{S}^\mathcal{A} \\ \boldsymbol{\gamma}_\mathcal{S}^\top \end{bmatrix} \boldsymbol{\zeta}_\mathcal{S} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\zeta}_{\mathcal{S}^c} = \mathbf{0}, \tag{4.6}$$

then it holds for arbitrary $\alpha > 0$ that

$$\boldsymbol{\Phi}(\boldsymbol{z}^\ell + \alpha\boldsymbol{\zeta}) = \boldsymbol{u}, \quad \boldsymbol{\Psi}^\mathcal{A}(\boldsymbol{z}^\ell + \alpha\boldsymbol{\zeta}) = \boldsymbol{v}_\mathcal{A} \quad \text{and} \quad \boldsymbol{z}_{\mathcal{S}^c}^\ell + \alpha\boldsymbol{\zeta}_{\mathcal{S}^c} = \mathbf{0}. \tag{4.7}$$

The largest $\alpha > 0$ such that also

$$\boldsymbol{\Psi}^{\mathcal{A}^c}(\boldsymbol{z}^\ell + \alpha\boldsymbol{\zeta}) \geq \boldsymbol{v}_{\mathcal{A}^c} \quad \text{and} \quad \operatorname{Diag}(\boldsymbol{\sigma}_\mathcal{S})(\boldsymbol{z}_\mathcal{S}^\ell + \alpha\boldsymbol{\zeta}_\mathcal{S}) \geq \mathbf{0} \tag{4.8}$$

is given by

$$\alpha = \min\left(\min_{\substack{i \in \mathcal{A}^c \\ \boldsymbol{\psi}_i^\top \boldsymbol{\zeta} < 0}} \frac{v_i - \boldsymbol{\psi}_i^\top \boldsymbol{z}^\ell}{\boldsymbol{\psi}_i^\top \boldsymbol{\zeta}}, \min_{\substack{j \in \mathcal{S} \\ \sigma_j \zeta_j < 0}} -\frac{z_j^\ell}{\zeta_j} \right). \tag{4.9}$$

Note that $0 \leq \alpha < \infty$ since we assumed that (4.1) is bounded. The sets

$$\mathcal{A}^+ = \left\{ i \in \mathcal{A}^c : \boldsymbol{\psi}_i^\top(\boldsymbol{z}^\ell + \alpha\boldsymbol{\zeta}) = v_i \right\} \quad \text{and} \quad \mathcal{S}^- = \left\{ j \in \mathcal{S} : z_j^\ell + \alpha\zeta_j = 0 \right\} \tag{4.10}$$

are the index sets where the minimum is attained, i.e., the sets of *blocking constraints*. Each $i \in \mathcal{A}^+$ joins the active set and each $j \in \mathcal{S}^-$ leaves the support if we perform the step $\alpha\boldsymbol{\zeta}$. Consequently, we update $\boldsymbol{z}^{\ell+1} = \boldsymbol{z}^\ell + \alpha\boldsymbol{\zeta}$, $\mathcal{A} = \mathcal{A} \cup \mathcal{A}^+$ and $\mathcal{S} = \mathcal{S} \setminus \mathcal{S}^-$.

4.1.4 Lagrange Multipliers

If there is no direction according to (4.6), then zero is an optimal solution of

$$\begin{aligned} \min_{\boldsymbol{\zeta}_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}} \quad & \boldsymbol{\gamma}_\mathcal{S}^\top \boldsymbol{\zeta}_\mathcal{S} \\ \text{s.t.} \quad & \begin{bmatrix} \boldsymbol{\Phi}_\mathcal{S} \\ \boldsymbol{\Psi}_\mathcal{S}^\mathcal{A} \end{bmatrix} \boldsymbol{\zeta}_\mathcal{S} = \mathbf{0}. \end{aligned} \tag{4.11}$$

Employing KKT conditions again, we see that there exist $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}_{\mathcal{A}}$ satisfying (4.4a). For the case that $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}_{\mathcal{A}}$ additionally satisfy (4.4b) and (4.4c), Lemma 44 states that $\boldsymbol{z}^{\ell}$ is an optimal solution.

Otherwise, with $\boldsymbol{\nu}_{\mathcal{S}^c}$ according to (4.5), there exists at least one index $i \in \mathcal{A}$ such that $\mu_i < 0$ or $j \in \mathcal{S}^c$ such that $\nu_j < 0$. We select the smaller of both values and set $\mathcal{A} = \mathcal{A} \setminus \{i\}$ or $\mathcal{S} = \mathcal{S} \cup \{j\}$, respectively. Then, we search a new direction according to Subsection 4.1.3.

4.1.5 Feasibility of Generated Directions

In the context of the previous section, suppose that $\mu_i < 0$ and we set $\mathcal{A} = \mathcal{A} \setminus \{i\}$. Afterwards, we go back to (4.6) and find a new direction $\boldsymbol{\zeta}$. It holds that

$$\begin{aligned} -1 \overset{(4.6)}{=} \boldsymbol{\gamma}_{\mathcal{S}}^{\top}\boldsymbol{\zeta}_{\mathcal{S}} &\overset{(4.4a)}{=} (\boldsymbol{\Phi}_{\mathcal{S}}^{\top}\boldsymbol{\lambda} + (\boldsymbol{\Psi}_{\mathcal{S}}^{\mathcal{A}})^{\top}\boldsymbol{\mu}_{\mathcal{A}} + (\boldsymbol{\Psi}_{\mathcal{S}}^{i})^{\top}\mu_i)^{\top}\boldsymbol{\zeta}_{\mathcal{S}} \\ &= \boldsymbol{\lambda}^{\top}\boldsymbol{\Phi}_{\mathcal{S}}\boldsymbol{\zeta}_{\mathcal{S}} + \boldsymbol{\mu}_{\mathcal{A}}^{\top}\boldsymbol{\Psi}_{\mathcal{S}}^{\mathcal{A}}\boldsymbol{\zeta}_{\mathcal{S}} + \mu_i\boldsymbol{\Psi}_{\mathcal{S}}^{i}\boldsymbol{\zeta}_{\mathcal{S}} \\ &\overset{(4.6)}{=} \mu_i\boldsymbol{\psi}_i^{\top}\boldsymbol{\zeta}. \end{aligned} \tag{4.12}$$

It follows that $\boldsymbol{\psi}_i^{\top}\boldsymbol{\zeta} = -\mu_i^{-1} > 0$. Consequently, it holds that $\boldsymbol{\psi}_i^{\top}(\boldsymbol{z}^{\ell} + \alpha\boldsymbol{\zeta}) > v_i$ and the step $\alpha\boldsymbol{\zeta}$ preserves the property of $\mathcal{A}$ exactly reflecting the set of active constraints. An analogous statement holds if we update $\mathcal{S} = \mathcal{S} \cup \{j\}$ prior to finding a direction $\boldsymbol{\zeta}$. In this case, we obtain $\sigma_j\zeta_j = -\nu_j^{-1} > 0$.

4.1.6 Algorithm and Set Management

The steps that we have discussed in the previous subsections form the basis of the iterative scheme that is formalized in Algorithm 2. In short, the scheme can be summarized as follows: If there exists a descent direction $\boldsymbol{\zeta}$ in Step 5, we continue by calculating the associated step size α, perform the step $\alpha\boldsymbol{\zeta}$ and update the active set $\mathcal{A}$ as well as the support $\mathcal{S}$. In case a descent direction does not exist, we determine Lagrange multipliers $\boldsymbol{\mu}_{\mathcal{A}}$ and $\boldsymbol{\nu}_{\mathcal{S}^c}$. Either, these multipliers show that the current iterate $\boldsymbol{z}^{\ell}$ is already an optimal solution or they indicate which indices are to be removed from the active set or added to the support, respectively.

However, there are some special cases that need to be handled with care. First of all, even after $\mathcal{A}$ or $\mathcal{S}$ have been updated in Steps 22–32, it can occur that we do not find a new descent direction in Step 5. Consequently, either the active set or the support will be modified again. Suppose that the indices i_1 and i_2 were consecutively removed from the active set. Then, it does not necessarily hold for the next direction that $\boldsymbol{\psi}_{i_1}^{\top}\boldsymbol{\zeta} > 0$ and $\boldsymbol{\psi}_{i_2}^{\top}\boldsymbol{\zeta} > 0$ because if we proceed as in (4.12), we only obtain $\boldsymbol{\psi}_{i_1}^{\top}\boldsymbol{\zeta} + \boldsymbol{\psi}_{i_2}^{\top}\boldsymbol{\zeta} > 0$. An analogous statement holds in case one index is removed from the active set and one is added to the support, or if two indices are added to the support. Therefore, we keep track of all the indices that were consecutively removed from the active set as well as those indices that were consecutively added to the support via the sets $\mathcal{A}^{-}$ and $\mathcal{S}^{+}$.

Input: $\boldsymbol{\gamma} \in \mathbb{R}^d$, $\boldsymbol{\Phi} \in \mathbb{R}^{k \times d}$, $\boldsymbol{u} \in \mathbb{R}^k$, $\boldsymbol{\Psi} \in \mathbb{R}^{l \times d}$, $\boldsymbol{v} \in \mathbb{R}^l$, $\boldsymbol{\sigma} \in \{\pm 1\}^d$, feasible $\boldsymbol{z}^0 \in \mathbb{R}^d$ and associated sets $\mathcal{A}$ and $\mathcal{S}$
Output: solution $\boldsymbol{z}^\star$ to problem (4.1)

$\ell \leftarrow 0$
$\mathcal{A}^- \leftarrow \emptyset$
$\mathcal{S}^+ \leftarrow \emptyset$
while *not stopped* **do**
  if *a solution* $\boldsymbol{\zeta}$ *of* (4.6) *exists* **then**
    $\alpha \leftarrow$ step size according to (4.9)
    $\boldsymbol{z}^{\ell+1} \leftarrow \boldsymbol{z}^\ell + \alpha \boldsymbol{\zeta}$
    $(\mathcal{A}^+, \mathcal{S}^-) \leftarrow$ blocking constraints according to (4.10)
    $\mathcal{A} \leftarrow \mathcal{A} \cup \mathcal{A}^+$
    $\mathcal{S} \leftarrow \mathcal{S} \setminus \mathcal{S}^-$
    if $\alpha = 0$ **then**
      $\mathcal{A}^- \leftarrow \mathcal{A}^- \setminus \mathcal{A}^+$
      $\mathcal{S}^+ \leftarrow \mathcal{S}^+ \setminus \mathcal{S}^-$
    else
      if $|\mathcal{A}^-| + |\mathcal{S}^+| > 1$ **then**
        $\mathcal{A} \leftarrow \mathcal{A} \cup \left\{ i \in \mathcal{A}^- : \boldsymbol{\psi}_i^\top \boldsymbol{\zeta} = 0 \right\}$
        $\mathcal{S} \leftarrow \mathcal{S} \setminus \{ j \in \mathcal{S}^+ : \zeta_j = 0 \}$
      $\mathcal{A}^- \leftarrow \emptyset$
      $\mathcal{S}^+ \leftarrow \emptyset$
    $\ell \leftarrow \ell + 1$
  else
    $(\boldsymbol{\mu}_\mathcal{A}, \boldsymbol{\nu}_{\mathcal{S}^c}) \leftarrow$ Lagrange multipliers according to (4.4a) and (4.5)
    $i^- \leftarrow \arg\min_{i \in \mathcal{A}} \mu_i$
    $j^+ \leftarrow \arg\min_{j \in \mathcal{S}^c} \nu_j$
    if $\mu_{i^-} \geq 0$ *and* $\nu_{j^+} \geq 0$ **then**
      return $\boldsymbol{z}^\star = \boldsymbol{z}^\ell$
    else if $\mu_{i^-} < \nu_{j^+}$ **then**
      $\mathcal{A} \leftarrow \mathcal{A} \setminus \{i^-\}$
      $\mathcal{A}^- \leftarrow \mathcal{A}^- \cup \{i^-\}$
    else
      $\mathcal{S} \leftarrow \mathcal{S} \cup \{j^+\}$
      $\mathcal{S}^+ \leftarrow \mathcal{S}^+ \cup \{j^+\}$

Algorithm 2: Active-Set Method for LPs.

Now, suppose that we find a descent direction in Step 5 but afterwards obtain the step size $\alpha = 0$. The only chance for this to happen is when $|\mathcal{A}^-| + |\mathcal{S}^+| > 1$ and $\boldsymbol{\psi}_i^\top \boldsymbol{\zeta} < 0$ for some $i \in \mathcal{A}^-$ or $\sigma_j \zeta_j < 0$ for some $j \in \mathcal{S}^+$. Afterwards in Step 8, the sets $\mathcal{A}^+$ and $\mathcal{S}^-$ contain exactly all i and j that have the respective property. Since in the following Steps 9 and 10, we change the active set and the support, we modify $\mathcal{A}^-$ and $\mathcal{S}^+$ accordingly in Steps 12 and 13.

If, in contrast, we obtain $\alpha > 0$ in Step 6, it can still be true that $|\mathcal{A}^-| + |\mathcal{S}^+| > 1$ and $\boldsymbol{\psi}_i^\top \boldsymbol{\zeta} = 0$ for some $i \in \mathcal{A}^-$ or $\zeta_j = 0$ for some $j \in \mathcal{S}^+$. Then, the respective indices i and j need to be re-added to the active set or re-removed from the support in Steps 16 and 17, respectively. The fact that $i \in \mathcal{A}^-$ means that the i-th constraint could potentially have become inactive. However, the actual descent direction $\boldsymbol{\zeta}$ does not reflect this property. Analogously, the new iterate could potentially have become non-zero in the j-th component, which did not happen due to $\zeta_j = 0$.

4.2 Active-Set Method for the Dual Update

The task to determine a new dual iterate $\boldsymbol{y}^{k+1}$ as a minimizer of (U_D^k) gives rise to the linear program

$$\begin{aligned}
\min_{\boldsymbol{y}_W \in \mathbb{R}^{|W|}} \quad & -\operatorname{sign}(\boldsymbol{A}^W \boldsymbol{x}^k - \boldsymbol{b}_W)^\top \boldsymbol{y}_W && && \text{(4.13a)}\\
\text{s.t.} \quad & (-\boldsymbol{A}_S^W)^\top \boldsymbol{y}_W &&= \operatorname{sign}(\boldsymbol{x}_S^k) && \text{(4.13b)}\\
& \begin{bmatrix} (\boldsymbol{A}_{S^c}^W)^\top \\ -(\boldsymbol{A}_{S^c}^W)^\top \end{bmatrix} \boldsymbol{y}_W &&\geq -\mathbb{1} && \text{(4.13c)}\\
& \operatorname{Diag}[\operatorname{sign}(\boldsymbol{A}^W \boldsymbol{x}^k - \boldsymbol{b}_W)] \boldsymbol{y}_W &&\geq \boldsymbol{0} && \text{(4.13d)}
\end{aligned}$$

whose structure corresponds to the form (4.1) and hence, we can apply the active-set method from the previous section. The notation that we use here is adapted to our previous notation in respect of the dual update. To make the following steps comprehensible, Table 4.1 contains a summary of the most important ingredients in terms of Section 4.1. Further, for reasons of clarity, we use the notation $S = S_k$, $W = W_k$, $\Omega = \Omega_k$ and $\Sigma = \Sigma_k$ throughout this section.

4.2.1 Initialization

In the beginning, $\boldsymbol{y}_W^k$ is feasible since $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ is an optimal pair. We set $\ell = 0$ and choose $\hat{\boldsymbol{y}}_W^0 = \boldsymbol{y}_W^k$ as our starting point for the active-set method. In view of (4.13c), the set of active constraints at $\hat{\boldsymbol{y}}^0$ corresponds to $\Sigma \setminus S$ with either positive or negative sign and the initial support is Ω (if we use $\hat{\boldsymbol{y}}^\ell$ without subscript at some points in the following, then we imply $\hat{\boldsymbol{y}}_{\Omega^c}^\ell = \boldsymbol{0}$).

Table 4.1: Active-set nomenclature in terms of Section 4.1 specialized for the dual update.

$\mathcal{A}$	$\Sigma \setminus S$
$\mathcal{A}^c$	Σ^c
$\mathcal{S}$	Ω
$\mathcal{S}^c$	$W \setminus \Omega$
$\boldsymbol{\gamma}_{\mathcal{S}}$	$-\text{sign}(\boldsymbol{A}^{\Omega}\boldsymbol{x}^k - \boldsymbol{b}_{\Omega})$
$\boldsymbol{\gamma}_{\mathcal{S}^c}$	$-\text{sign}(\boldsymbol{A}^{W\setminus\Omega}\boldsymbol{x}^k - \boldsymbol{b}_{W\setminus\Omega})$
$\boldsymbol{\Phi}_{\mathcal{S}}$	$(-\boldsymbol{A}^{\Omega}_{S})^\top$
$\boldsymbol{\Phi}_{\mathcal{S}^c}$	$(-\boldsymbol{A}^{W\setminus\Omega}_{S})^\top$
$\boldsymbol{\Psi}^{\mathcal{A}}_{\mathcal{S}}$	$(\boldsymbol{A}^\top_{\Sigma\setminus S}\hat{\boldsymbol{y}}^\ell) \odot (-\boldsymbol{A}^{\Omega}_{\Sigma\setminus S})^\top$
$\boldsymbol{\Psi}^{\mathcal{A}}_{\mathcal{S}^c}$	$(\boldsymbol{A}^\top_{\Sigma\setminus S}\hat{\boldsymbol{y}}^\ell) \odot (-\boldsymbol{A}^{W\setminus\Omega}_{\Sigma\setminus S})^\top$
$\boldsymbol{\sigma}_{\mathcal{S}}$	$\text{sign}(\boldsymbol{A}^{\Omega}\boldsymbol{x}^k - \boldsymbol{b}_{\Omega})$
$\boldsymbol{\sigma}_{\mathcal{S}^c}$	$\text{sign}(\boldsymbol{A}^{W\setminus\Omega}\boldsymbol{x}^k - \boldsymbol{b}_{W\setminus\Omega})$

4.2.2 Descent Direction and Blocking Constraints

If we transfer (4.6) to the situation in the dual update and call the sought-after descent direction $\boldsymbol{e}$ (where we imply again that $\boldsymbol{e}_{\Omega^c} = \boldsymbol{0}$), then we arrive at the system

$$\begin{aligned} (\boldsymbol{A}^{\Omega}_{\Sigma})^\top \boldsymbol{e}_{\Omega} &= \boldsymbol{0} \\ \text{sign}(\boldsymbol{A}^{\Omega}\boldsymbol{x}^k - \boldsymbol{b}_{\Omega})^\top \boldsymbol{e}_{\Omega} &= 1 \end{aligned} \tag{4.14}$$

of $|\Sigma|+1$ linear equations in $|\Omega|$ variables. If such a direction exists, the largest feasibility-preserving step size is

$$\alpha = \min\left\{\alpha_{\Sigma^c}, \alpha_{\Omega}\right\}, \tag{4.15}$$

where

$$\alpha_{\Sigma^c} = \min\left\{ \min_{\substack{j\in\Sigma^c \\ \boldsymbol{A}_j^\top \boldsymbol{e}<0}} \frac{1 + \boldsymbol{A}_j^\top \hat{\boldsymbol{y}}^\ell}{-\boldsymbol{A}_j^\top \boldsymbol{e}}, \min_{\substack{j\in\Sigma^c \\ \boldsymbol{A}_j^\top \boldsymbol{e}>0}} \frac{1 - \boldsymbol{A}_j^\top \hat{\boldsymbol{y}}^\ell}{\boldsymbol{A}_j^\top \boldsymbol{e}} \right\} \quad \text{and} \tag{4.16}$$

$$\alpha_{\Omega} = \min_{\substack{i\in\Omega \\ \text{sign}(\boldsymbol{a}_i^\top \boldsymbol{x}^k - b_i)e_i<0}} -\frac{\hat{y}_i^\ell}{e_i}. \tag{4.17}$$

Therewith, the new iterate is $\hat{\boldsymbol{y}}^{\ell+1} = \hat{\boldsymbol{y}}^\ell + \alpha\boldsymbol{e}$ and we need to update

$$\begin{aligned} \Omega &= \Omega \setminus \left\{ i \in \Omega : \hat{y}_i^{\ell+1} = 0 \right\} \\ \Sigma &= \Sigma \cup \left\{ j \in \Sigma^c : |\boldsymbol{A}_j^\top \hat{\boldsymbol{y}}^{\ell+1}| = 1 \right\}. \end{aligned} \tag{4.18}$$

4.2.3 Lagrange Multipliers

In case there is no solution of (4.14), there exist Lagrange multipliers according to (4.4a). After an adequate substitution, the respective system adapted to the dual update can be written as

$$\boldsymbol{A}_{\Sigma}^{\Omega}\hat{\boldsymbol{d}}_{\Sigma} = -\operatorname{sign}(\boldsymbol{A}^{\Omega}\boldsymbol{x}^{k} - \boldsymbol{b}_{\Omega}). \tag{4.19}$$

The Lagrange multipliers associated with the active set are then

$$\boldsymbol{\mu}_{\Sigma\setminus S} = -(\boldsymbol{A}_{\Sigma\setminus S}^{\top}\hat{\boldsymbol{y}}^{\ell}) \odot \hat{\boldsymbol{d}}_{\Sigma\setminus S} \tag{4.20}$$

and the multipliers related to the support are

$$\boldsymbol{\nu}_{W\setminus\Omega} = -\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\boldsymbol{x}^{k} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}_{\Sigma}^{W\setminus\Omega}\hat{\boldsymbol{d}}_{\Sigma} - \mathbb{1}. \tag{4.21}$$

In case $\boldsymbol{\mu}_{\Sigma\setminus S} \geq 0$ and $\boldsymbol{\nu}_{W\setminus\Omega} \geq 0$, the current iterate $\hat{\boldsymbol{y}}^{\ell}$ is already an optimal solution of (4.13) and we return $\Omega_{k+1} = \Omega$ as well as $\Sigma_{k+1} = \Sigma$. Otherwise, we pick one $j \in \Sigma \setminus S$ with $\mu_j < 0$ or $i \in W \setminus \Omega$ with $\nu_j < 0$ and update $\Sigma = \Sigma \setminus \{j\}$ or $\Omega = \Omega \cup \{i\}$ accordingly.

4.3 Active-Set Method for the Primal Update

Finding a pair of a new primal iterate $\boldsymbol{x}^{k+1}$ and the related decrease of the homotopy parameter t^{k+1} as a minimizer of (U_P^k) gives rise to the linear program

$$\min_{(\boldsymbol{x}_\Sigma, t)\in\mathbb{R}^{|\Sigma|}\times\mathbb{R}} \quad \begin{pmatrix}\boldsymbol{0}\\ -1\end{pmatrix}^{\top}\begin{pmatrix}\boldsymbol{x}_\Sigma\\ t\end{pmatrix} \tag{4.22a}$$

$$\text{s.t.} \quad \begin{bmatrix}\boldsymbol{A}_{\Sigma}^{\Omega} & \operatorname{sign}(\boldsymbol{y}_{\Omega}^{k+1})\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_\Sigma\\ t\end{pmatrix} = \delta^{k}\operatorname{sign}(\boldsymbol{y}_{\Omega}^{k+1}) + \boldsymbol{b}_{\Omega} \tag{4.22b}$$

$$\begin{bmatrix}\boldsymbol{A}_{\Sigma}^{\Omega^c} & -\mathbb{1}\\ -\boldsymbol{A}_{\Sigma}^{\Omega^c} & -\mathbb{1}\\ \boldsymbol{0} & -1\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_\Sigma\\ t\end{pmatrix} \geq \begin{pmatrix}-\delta^{k}\mathbb{1} + \boldsymbol{b}_{\Omega^c}\\ -\delta^{k}\mathbb{1} - \boldsymbol{b}_{\Omega^c}\\ \delta - \delta^{k}\end{pmatrix} \tag{4.22c}$$

$$\begin{bmatrix}-\operatorname{Diag}(\boldsymbol{A}_{\Sigma}^{\top}\boldsymbol{y}^{k+1}) & 0\\ \boldsymbol{0} & 1\end{bmatrix}\begin{pmatrix}\boldsymbol{x}_\Sigma\\ t\end{pmatrix} \geq \boldsymbol{0}. \tag{4.22d}$$

Compared to the original formulation, we transformed the maximization problem into a minimization problem and flipped the relations in the inequality constraints in order to match the problem to the form (4.1). As in the previous section, we simplify our notation by writing $S = S_k$, $W = W_k$, $\Omega = \Omega_{k+1}$ and $\Sigma = \Sigma_{k+1}$. Also, we summarize the essential elements of the active-set method for the primal update in Table 4.2.

4.3.1 Initialization

The point $(\boldsymbol{x}_{\Sigma}^{k}, 0)$ is feasible for (4.22) since $(\boldsymbol{x}^{k}, \boldsymbol{y}^{k+1})$ is an optimal pair for (P_{δ^k}). Accordingly, we set $\ell = 0$ and choose our starting point to be $(\hat{\boldsymbol{x}}_{\Sigma}^{0}, \tau^{0}) = (\boldsymbol{x}_{\Sigma}^{k}, 0)$. Regarding (4.22c), we see that the subset of active constraints at this point corresponds

Table 4.2: Active-set nomenclature in terms of Section 4.1 specialized for the primal update.

$\mathcal{A}$	$W \setminus \Omega$
$\mathcal{A}^c$	$W^c \cup \{m+1\}$
$\mathcal{S}$	$S \cup \{t\}$
$\mathcal{S}^c$	$\Sigma \setminus S$
$\boldsymbol{\gamma}_{\mathcal{S}}$	$(\mathbf{0};\ -1)$
$\boldsymbol{\gamma}_{\mathcal{S}^c}$	$\mathbf{0}$
$\boldsymbol{\Phi}_{\mathcal{S}}$	$\begin{bmatrix} \boldsymbol{A}_S^{\Omega} & \operatorname{sign}(\boldsymbol{y}_{\Omega}^{k+1}) \end{bmatrix}$
$\boldsymbol{\Phi}_{\mathcal{S}^c}$	$\boldsymbol{A}_{\Sigma\setminus S}^{\Omega}$
$\boldsymbol{\Psi}_{\mathcal{S}}^{\mathcal{A}}$	$\begin{bmatrix} -\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}}^{\ell} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}_S^{W\setminus\Omega} & -\mathbb{1} \end{bmatrix}$
$\boldsymbol{\Psi}_{\mathcal{S}^c}^{\mathcal{A}}$	$-\operatorname{sign}(\boldsymbol{A}^{W\setminus\Omega}\hat{\boldsymbol{x}}^{\ell} - \boldsymbol{b}_{W\setminus\Omega}) \odot \boldsymbol{A}_{\Sigma\setminus S}^{W\setminus\Omega}$
$\boldsymbol{\sigma}_{\mathcal{S}}$	$(\boldsymbol{A}_S^{\top}\boldsymbol{y}^{k+1};\ 1)$
$\boldsymbol{\sigma}_{\mathcal{S}^c}$	$\boldsymbol{A}_{\Sigma\setminus S}^{\top}\boldsymbol{y}^{k+1}$

to $W\setminus\Omega$ with either positive or negative sign. Further, the initial support is S (analogous to above, we imply $\hat{\boldsymbol{x}}_{S^c} = 0$ in case we omit the subscript).

The variable t represents the decrease of the homotopy parameter starting from δ^k. Although the associated iterate is initially zero, it is clear from the objective function in (4.22a) that t must join the support before we can make a step towards a descent direction. As this is known in advance, we can forgo the calculation of Lagrange multipliers in the first iteration and directly initialize the support as $S \cup \{t\}$. Note that the constraint $t \geq 0$ in (4.22d) can actually be omitted. Nevertheless, we keep it in order to adapt (4.22) to (4.1). Moreover, the constraint $-t \geq \delta - \delta^k$ in (4.22c) is neither active in the beginning nor will it become so unless we have found an optimal solution of our original problem (P_δ). In Table 4.2, we refer to this constraint using the index $m+1$.

4.3.2 Descent Direction and Blocking Constraints

In the following, the notation $(\boldsymbol{d}, d_t)$ refers to a descent direction with respect to the current iterate $(\hat{\boldsymbol{x}}^{\ell}, \tau^{\ell})$, where $\boldsymbol{d}_{S^c} = \mathbf{0}$. Due to the simple structure of the objective function in (4.22a), where the only non-zero coefficient relates to t, each associated descent direction according to (4.6) satisfies $d_t = 1$. If we fix this component in advance and moreover use that $\operatorname{sign}(\boldsymbol{y}_{\Omega}^{k+1}) = \operatorname{sign}(\boldsymbol{A}^{\Omega}\hat{\boldsymbol{x}}^{\ell} - \boldsymbol{b}_{\Omega})$, then we need to solve the linear equation system

$$\boldsymbol{A}_S^W \boldsymbol{d}_S = -\operatorname{sign}(\boldsymbol{A}^W \hat{\boldsymbol{x}}^{\ell} - \boldsymbol{b}_W) \tag{4.23}$$

with $|W|$ constraints and $|S|$ variables in order to obtain the full descent direction $(\boldsymbol{d}, 1)$. In case such a direction exists, the step size according to (4.9) is

$$\alpha = \min\left\{\alpha_{W^c}, \alpha_S, \delta^k - \tau^\ell - \delta\right\}, \tag{4.24}$$

wherein

$$\alpha_{W^c} = \min\left\{ \min_{\substack{i \in W^c \\ \boldsymbol{a}_i^\top \boldsymbol{d} > -1}} \frac{\delta^k - \tau^\ell - \boldsymbol{a}_i^\top \hat{\boldsymbol{x}}^\ell + b_i}{\boldsymbol{a}_i^\top \boldsymbol{d} + 1}, \min_{\substack{i \in W^c \\ \boldsymbol{a}_i^\top \boldsymbol{d} < 1}} \frac{\delta^k - \tau^\ell + \boldsymbol{a}_i^\top \hat{\boldsymbol{x}}^\ell - b_i}{-\boldsymbol{a}_i^\top \boldsymbol{d} + 1} \right\} \tag{4.25}$$

and

$$\alpha_S = \min_{\substack{j \in S \\ \boldsymbol{A}_j^\top \boldsymbol{y}^{k+1} \cdot d_j > 0}} -\frac{\hat{x}_j^\ell}{d_j}. \tag{4.26}$$

The new iterates are then $\hat{\boldsymbol{x}}^{\ell+1} = \hat{\boldsymbol{x}}^\ell + \alpha \boldsymbol{d}$ and $\tau^{\ell+1} = \tau^\ell + \alpha$. In case $\alpha = \delta^k - \tau^\ell - \delta$, we stop thereafter since $\boldsymbol{x}^\star = \hat{\boldsymbol{x}}^{\ell+1}$ is an optimal solution of (P_δ). Otherwise, we finally update

$$\begin{aligned} W &= W \cup \left\{ i \in W^c : |\boldsymbol{a}_i^\top \hat{\boldsymbol{x}}^{\ell+1} - b_i| = \delta^k - \tau^{\ell+1} \right\} \\ S &= S \setminus \left\{ j \in S : \hat{x}_j^{\ell+1} = 0 \right\}. \end{aligned} \tag{4.27}$$

4.3.3 Lagrange Multipliers

If there is no direction satisfying (4.23), then there exist Lagrange multipliers according to (4.4a). Performing a suitable substitution and using once more that $\operatorname{sign}(\boldsymbol{y}_\Omega^{k+1}) = \operatorname{sign}(\boldsymbol{A}^\Omega \hat{\boldsymbol{x}}^\ell - \boldsymbol{b}_\Omega)$, we arrive at the system

$$\begin{aligned} (\boldsymbol{A}_S^W)^\top \hat{\boldsymbol{e}}_W &= \boldsymbol{0} \\ \operatorname{sign}(\boldsymbol{A}^W \hat{\boldsymbol{x}}^\ell - \boldsymbol{b}_W)^\top \hat{\boldsymbol{e}}_W &= 1. \end{aligned} \tag{4.28}$$

The solution then leads us to the Lagrange multipliers

$$\boldsymbol{\mu}_{W \setminus \Omega} = \operatorname{sign}(\boldsymbol{A}^{W \setminus \Omega} \hat{\boldsymbol{x}}^\ell - \boldsymbol{b}_{W \setminus \Omega}) \odot \hat{\boldsymbol{e}}_{W \setminus \Omega} \tag{4.29}$$

and

$$\boldsymbol{\nu}_{\Sigma \setminus S} = -(\boldsymbol{A}_{\Sigma \setminus S}^\top \boldsymbol{y}^{k+1}) \odot (\boldsymbol{A}_{\Sigma \setminus S}^W)^\top \hat{\boldsymbol{e}}_W \tag{4.30}$$

associated to the active set and the support, respectively. In case both $\boldsymbol{\mu}_{W \setminus \Omega} \geq 0$ and $\boldsymbol{\nu}_{\Sigma \setminus S} \geq 0$, the current iterate $(\hat{\boldsymbol{x}}^\ell, \tau^\ell)$ is an optimal solution of (4.22). If not, there exists an $i \in W \setminus \Omega$ with $\mu_i < 0$ or a $j \in \Sigma \setminus S$ with $\nu_j < 0$ and we update $W = W \setminus \{i\}$ or $S = S \cup \{j\}$, respectively.

4.4 The Ambiguity of Lagrange Multipliers

The active-set methods which we developed in the previous two sections have one structural aspect in common: As long as we update the dual iterate, the primal support S

and the primal active set W remain unchanged. Vice versa, the dual support Ω and the dual active set Σ are not modified during the update of the primal iterate. Now, imagine the situation at the very beginning of a dual update step, say we have an optimal pair $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ and seek for $\boldsymbol{y}^{k+1}$, where $\ell = 0$ and we try to find a direction $\boldsymbol{e}$ according to (4.14) for the first time. Since $\boldsymbol{A}$ and $\boldsymbol{b}$ are naturally invariant throughout the algorithm, the linear equation system (4.14) depends solely on Ω and Σ. These two sets did not change after the previous dual update in which we identified $\boldsymbol{y}^k$. The point is now that we have already tried to find a descent direction at $\boldsymbol{y}^k$ with exactly the same Ω and Σ before the end of the previous dual update, which did however not exist (after that, we identified Lagrange multipliers which certified the optimality of $\boldsymbol{y}^k$). Hence, we have no chance to find a descent direction at the beginning of the dual update for $\boldsymbol{y}^{k+1}$ without first changing Ω or Σ via Lagrange multipliers (which are different from the multipliers that certified the optimality of $\boldsymbol{y}^k$ as the sets S and W have changed in the meantime).

The same reasoning as above applies in terms of a primal update step. In both cases we are not able to identify a descent direction at the beginning of the respective active-set method because the respective support and active set need to be modified first via Lagrange multipliers. In the following, we pursue an elegant way to avoid the calculation of Lagrange multipliers in the first iteration of our active-set methods and even to derive an initial descent direction from a previous update. It will turn out that the solution $\hat{\boldsymbol{e}}_W$ of (4.28) inducing Lagrange multipliers $\boldsymbol{\mu}_{W\setminus\Omega} \geq \boldsymbol{0}$ and $\boldsymbol{\nu}_{\Sigma\setminus S} \geq \boldsymbol{0}$ at the end of the primal update for $\boldsymbol{x}^k$ is also a suitable first descent direction in the dual update for $\boldsymbol{y}^{k+1}$. Analogously, the solution $\hat{\boldsymbol{d}}_\Sigma$ of (4.19) that leads to Lagrange multipliers $\boldsymbol{\mu}_{\Sigma\setminus S} \geq \boldsymbol{0}$ and $\boldsymbol{\nu}_{W\setminus\Sigma} \geq \boldsymbol{0}$ showing the optimality of $\boldsymbol{y}^{k+1}$, is an appropriate descent direction at the beginning of the primal update for $\boldsymbol{x}^{k+1}$. The *ambiguity* of Lagrange multipliers that the title of this section suggests refers to these circumstances.

4.4.1 Initial Direction for the Dual Update

Suppose again that we have an optimal $(\boldsymbol{x}^k, \boldsymbol{y}^k)$ at hand and find ourselves at the beginning of the dual update for $\boldsymbol{y}^{k+1}$ and let $\hat{\boldsymbol{e}}_W$ be the solution of (4.28) at the end of the primal update for $\boldsymbol{x}^k$. A comparison of (4.28) and the conditions (4.14) for a descent direction $\boldsymbol{e}_\Omega$ reveals that the associated linear equation systems are

$$\begin{aligned} (\boldsymbol{A}_S^W)^\top \hat{\boldsymbol{e}}_W &= \boldsymbol{0}, \\ \operatorname{sign}(\boldsymbol{A}^W \hat{\boldsymbol{x}}^\ell - \boldsymbol{b}_W)^\top \hat{\boldsymbol{e}}_W &= 1, \end{aligned} \quad \text{and} \quad \begin{aligned} (\boldsymbol{A}_\Sigma^\Omega)^\top \boldsymbol{e}_\Omega &= \boldsymbol{0}, \\ \operatorname{sign}(\boldsymbol{A}^\Omega \boldsymbol{x}^k - \boldsymbol{b}_\Omega)^\top \boldsymbol{e}_\Omega &= 1. \end{aligned} \tag{4.31}$$

As we have argued above, a solution to the system on the right-hand side does not exist unless we update the sets Ω and Σ. We recall that $\Omega \subseteq W$ and $S \subseteq \Sigma$ and hence, if we perform the updates

$$\begin{aligned} \Omega &\leftarrow \Omega \cup \{i \in W \setminus \Omega : \hat{e}_i \neq 0\}, \\ \Sigma &\leftarrow \Sigma \setminus \{j \in \Sigma \setminus S : (\boldsymbol{A}_j^W)^\top \hat{\boldsymbol{e}}_W \neq 0\}, \end{aligned} \tag{4.32}$$

then $\boldsymbol{e}_\Omega = \hat{\boldsymbol{e}}_\Omega$ is a solution of the system on the right-hand side. It remains to show that a non-trivial step $\boldsymbol{y}_\Omega^k + \alpha\hat{\boldsymbol{e}}_\Omega$ maintains primal-dual optimality. To that end, consider the Lagrange multipliers $\boldsymbol{\mu}_{W\setminus\Omega} \geq \mathbf{0}$ and $\boldsymbol{\nu}_{\Sigma\setminus S} \geq \mathbf{0}$ that are associated with $\hat{\boldsymbol{e}}_W$. For each $i \in W \setminus \Omega$ with $\hat{e}_i \neq 0$, it holds that $\mu_i = \operatorname{sign}(\boldsymbol{a}_i^\top \hat{\boldsymbol{x}}^k - b_i)\hat{e}_i > 0$, which shows that the respective components of the dual variable are provided with the correct sign. Furthermore, it holds for each $j \in \Sigma \setminus S$ with $(\boldsymbol{A}_j^W)^\top \hat{\boldsymbol{e}}_W \neq 0$ that $\boldsymbol{A}_j^\top \boldsymbol{y}^k \cdot (\boldsymbol{A}_j^W)^\top \hat{\boldsymbol{e}}_W < 0$, which shows that with a step in direction $\hat{\boldsymbol{e}}$, the respective dual constraint becomes inactive while feasibility is maintained.

4.4.2 Initial Direction for the Primal Update

The above idea to find an initial direction for the dual update can easiliy be transferred to the primal update. To that end, suppose that $(\boldsymbol{x}^k, \boldsymbol{y}^{k+1})$ is an optimal pair and we seek to find the next primal iterate $\boldsymbol{x}^{k+1}$, and let $\hat{\boldsymbol{d}}_\Sigma$ be the solution of (4.19) at the end of the previous dual update. The linear equation systems associated to (4.19) as well as (4.23) are

$$\boldsymbol{A}_\Sigma^\Omega \hat{\boldsymbol{d}}_\Sigma = -\operatorname{sign}(\boldsymbol{A}^\Omega \boldsymbol{x}^k - \boldsymbol{b}_\Omega) \qquad \text{and} \qquad \boldsymbol{A}_S^W \boldsymbol{d}_S = -\operatorname{sign}(\boldsymbol{A}^W \hat{\boldsymbol{x}}^k - \boldsymbol{b}_W), \tag{4.33}$$

respectively, where $\boldsymbol{d}_S$ is the sought-after descent direction starting from $\boldsymbol{x}^k$ and the associated system on the right-hand side does not have a solution before an update of S and W. Because $S \subseteq \Sigma$ and $\Omega \subseteq W$, we can update both sets according to

$$\begin{aligned} S &\leftarrow S \cup \{j \in \Sigma \setminus S : \hat{d}_j \neq 0\}, \\ W &\leftarrow W \setminus \{i \in W \setminus \Omega : \boldsymbol{A}_\Sigma^i \hat{\boldsymbol{d}}_\Sigma \neq -\operatorname{sign}(\boldsymbol{A}^i \boldsymbol{x}^k - b_i)\}, \end{aligned} \tag{4.34}$$

and see that $\boldsymbol{d}_S = \hat{\boldsymbol{d}}_S$ is then a solution of (4.23). Finally, we need to show that a step $\boldsymbol{x}_S^k + \alpha\hat{\boldsymbol{d}}_S$ maintains primal-dual optimality. For that purpose, we consider the Lagrange multipliers $\boldsymbol{\mu}_{\Sigma\setminus S} \geq \mathbf{0}$ and $\boldsymbol{\nu}_{W\setminus\Omega} \geq \mathbf{0}$ that come along with $\hat{\boldsymbol{d}}_\Sigma$. Each $j \in \Sigma \setminus S$ with $\hat{d}_j \neq 0$ satisfies $\mu_j = -\boldsymbol{A}_j^\top \boldsymbol{y}^{k+1} \cdot \hat{d}_j > 0$ and hence, the respective component of the primal iterate comes with the right sign. Last but not least, it holds for each $i \in W \setminus \Omega$ with $\boldsymbol{A}_\Sigma^i \hat{\boldsymbol{d}}_\Sigma \neq -\operatorname{sign}(\boldsymbol{A}^i \boldsymbol{x}^k - b_i)$ that $\nu_j = -\operatorname{sign}(\boldsymbol{A}^i \boldsymbol{x}^k - b_i) \cdot \boldsymbol{A}_\Sigma^i \hat{\boldsymbol{d}}_\Sigma - 1 > 0$ which shows that a step in the respective direction preserves feasibility and causes the i-th constraint to become inactive.

5 Connection to Linear Programming and Extensions

Both the ℓ_1-norm objective function and the ℓ_∞-norm constraint give the problem (P_δ) a clearly non-linear character. Nevertheless, it is possible to reformulate the problem as a linear program by applying well-known techniques. One the one hand, it is obvious that the ℓ_∞-norm constraint can be formulated in a linear fashion as

$$-\delta\mathbb{1} \leq \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \delta\mathbb{1} \;\Leftrightarrow\; \begin{bmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \end{bmatrix} \boldsymbol{x} \leq \begin{pmatrix} \delta\mathbb{1} + \boldsymbol{b} \\ \delta\mathbb{1} - \boldsymbol{b} \end{pmatrix}. \tag{5.1}$$

On the other hand, we show in Section 5.1 that the ℓ_1-norm objective function can be rewritten as a linear function if we use *variable splitting*.

5.1 Associated Linear Programs

We define the *positive part* and the *negative part* of $\boldsymbol{x}$ as two vectors $\boldsymbol{x}^+ \geq \boldsymbol{0}$ and $\boldsymbol{x}^- \geq \boldsymbol{0}$, respectively, such that $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-$ and $\boldsymbol{x}^+ \odot \boldsymbol{x}^- = \boldsymbol{0}$. The latter condition has the effect that at most one of both vectors can be non-zero in each component at a time. Hence, it holds that $\|\boldsymbol{x}\|_1 = \mathbb{1}^\top(\boldsymbol{x}^+ + \boldsymbol{x}^-)$. The following lemma exploits this relation and shows that both (P_δ) as well as (D_δ) have corresponding linear programs which are again dual to each other.

Lemma 45. *A vector $\boldsymbol{x}^\star$ is an optimal solution of (P_δ) if and only if its positive and negative part form an optimal solution of the linear program*

$$\min_{\boldsymbol{x}^\pm \in \mathbb{R}^n} \; \mathbb{1}^\top \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} \quad \text{s.t.} \quad \begin{bmatrix} -\boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & -\boldsymbol{A} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} \geq \begin{pmatrix} -\boldsymbol{b} - \delta\mathbb{1} \\ \boldsymbol{b} - \delta\mathbb{1} \end{pmatrix} \quad \boldsymbol{x}^\pm \geq \boldsymbol{0}. \tag{5.2}$$

Moreover, $\boldsymbol{y}^\star$ is an optimal solution of (D_δ) if and only if its positive and negative part form an optimal solution of the linear program

$$\max_{y^\pm \in \mathbb{R}^m} \; \begin{pmatrix} -\boldsymbol{b} - \delta\mathbb{1} \\ \boldsymbol{b} - \delta\mathbb{1} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{y}^+ \\ \boldsymbol{y}^- \end{pmatrix} \quad \text{s.t.} \quad \begin{bmatrix} -\boldsymbol{A}^\top & \boldsymbol{A}^\top \\ \boldsymbol{A}^\top & -\boldsymbol{A}^\top \end{bmatrix} \begin{pmatrix} \boldsymbol{y}^+ \\ \boldsymbol{y}^- \end{pmatrix} \leq \mathbb{1} \quad \boldsymbol{y}^\pm \geq \boldsymbol{0} \tag{5.3}$$

which is, in particular, the dual linear program associated with (5.2).

Proof. The fact that the problems (5.2) and (5.3) are connected via duality follows directly from the definition of the dual linear program (see [29], among many others). By complementary slackness (see [29, Theorem 2]), it holds that $\boldsymbol{x}^{\pm}$ and $\boldsymbol{y}^{\pm}$ are optimal solutions of (5.2) and (5.3), respectively, if and only if

$$\begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} \odot \left(\begin{bmatrix} -\boldsymbol{A}^\top & \boldsymbol{A}^\top \\ \boldsymbol{A}^\top & -\boldsymbol{A}^\top \end{bmatrix} \begin{pmatrix} \boldsymbol{y}^+ \\ \boldsymbol{y}^- \end{pmatrix} - \mathbb{1} \right) = \boldsymbol{0}, \tag{5.4}$$

$$\begin{pmatrix} \boldsymbol{y}^+ \\ \boldsymbol{y}^- \end{pmatrix} \odot \left(\begin{bmatrix} -\boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & -\boldsymbol{A} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} - \begin{pmatrix} -\boldsymbol{b} - \delta\mathbb{1} \\ \boldsymbol{b} - \delta\mathbb{1} \end{pmatrix} \right) = \boldsymbol{0} \tag{5.5}$$

and both are feasible solutions of the respective problems. If $x_j^+ > 0$, then it follows from (5.4) that $-\boldsymbol{A}_j^\top(\boldsymbol{y}^+ - \boldsymbol{y}^-) = 1$. Vice versa, it is clear that $\boldsymbol{A}_j^\top(\boldsymbol{y}^+ - \boldsymbol{y}^-) \neq 1$ and we conclude from (5.4) that $x_j^- = 0$. In the same way, we obtain that $x_j^+ = 0$ in case $x_j^- > 0$. Consequently, it holds that $\boldsymbol{x}^+ \odot \boldsymbol{x}^- = \boldsymbol{0}$ which shows that both vectors constitute the positive and negative part, respectively, of $\boldsymbol{x}^+ - \boldsymbol{x}^-$ and that the objective function of (5.2) can be rewritten as $\|\boldsymbol{x}^+ - \boldsymbol{x}^-\|_1$ at any optimal point. Analogously, by using (5.5), it follows that $\boldsymbol{y}^+ \odot \boldsymbol{y}^- = \boldsymbol{0}$ and that the objective function of (5.3) can be rewritten as $-\boldsymbol{b}^\top(\boldsymbol{y}^+ - \boldsymbol{y}^-) - \delta\|\boldsymbol{y}^+ - \boldsymbol{y}^-\|_1$ at any optimal point.

In view of (5.3), we can now conclude that it is equivalent to minimize the objective function $\|\boldsymbol{x}^+ - \boldsymbol{x}^-\|_1$ if we add the additional constraint $\boldsymbol{x}^+ \odot \boldsymbol{x}^- = \boldsymbol{0}$. As there is a one-to-one correspondence between $\mathbb{R}^n$ and the set $\{\boldsymbol{x}^+ - \boldsymbol{x}^- : \boldsymbol{x}^\pm \geq \boldsymbol{0} \text{ and } \boldsymbol{x}^+ \odot \boldsymbol{x}^- = \boldsymbol{0}\}$, it is further equivalent to substitute $\boldsymbol{x} \coloneqq \boldsymbol{x}^+ - \boldsymbol{x}^-$ and drop the constraints $\boldsymbol{x}^\pm \geq \boldsymbol{0}$ and $\boldsymbol{x}^+ \odot \boldsymbol{x}^- = \boldsymbol{0}$. A look at (5.1) finally reveals that

$$\begin{bmatrix} -\boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & -\boldsymbol{A} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} \geq \begin{pmatrix} -\boldsymbol{b} - \delta\mathbb{1} \\ \boldsymbol{b} - \delta\mathbb{1} \end{pmatrix} \Leftrightarrow \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \delta. \tag{5.6}$$

All in all, it follows that the split into the positive and negative parts of a vector establishes an exact correspondence between the minimizers of (5.2) and (P_δ) and that both problems have the same optimal value. Completely analogously, it can be seen that the statement about (5.3) and (D_δ) is true as well. □

5.2 Parametric Simplex Method

In the previous section, we have seen that the problem (P_δ) can be recast as an equivalent LP in the form of (5.2). Introducing non-negative *slack variables* $\boldsymbol{s}^\pm \in \mathbb{R}^m$, this formulation can be extended straightforwardly to the equality constrained LP

$$\begin{aligned} \min_{\boldsymbol{x}^\pm \in \mathbb{R}^n, \boldsymbol{s}^\pm \in \mathbb{R}^m} \quad & \mathbb{1}^\top \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} \quad \text{s.t.} \quad \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{A} & \boldsymbol{I}_m & \boldsymbol{0} \\ -\boldsymbol{A} & \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{I}_m \end{bmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \\ \boldsymbol{s}^+ \\ \boldsymbol{s}^- \end{pmatrix} = \begin{pmatrix} \boldsymbol{b} + \delta\mathbb{1} \\ -\boldsymbol{b} + \delta\mathbb{1} \end{pmatrix} \\ & \qquad\qquad \boldsymbol{x}^\pm \geq \boldsymbol{0} \\ & \qquad\qquad \boldsymbol{s}^\pm \geq \boldsymbol{0}. \end{aligned} \tag{5.7}$$

The LP homotopy method most naturally related to our approach results from treating the parameter δ as the homotopy parameter (as we also do in our method) in the equality constrained LP (5.7)—the so-called (self-dual) *parametric simplex method* (PSM) [15, 44]. Very briefly, PSM perturbs both the LP right-hand side and objective coefficient vectors using the same parameter and then drives this parameter down to zero, performing primal or dual simplex pivot steps at each breakpoint in the (piecewise linear) parameter homotopy path. Analogous to our approach, a primal-dual feasible (hence, optimal) basis is easily found and used to start the algorithm. Reducing the parameter, basis optimality is maintained until either a basic variable or nonbasic reduced cost coefficient changes sign, which identifies the breakpoints and induces an appropriate simplex step to exchange some basis element for a nonbasic one. For a detailed formal description, we refer to [44, pp. 115–121].

In fact, PSM was very recently proposed for sparse linear discriminant analysis problems by means of reformulating the associated problem (P_δ) (cf. [7] and Section 6.4) as precisely the LP stated above (see [37], where PSM is applied to several other problems as well). For the above special parameterized LP, one needs to stop PSM as soon as the parameter drops below the target original δ (*not* zero) and since the objective is unperturbed, only primal simplex pivot steps are performed throughout the entire algorithmic process (i.e., each breakpoint identifies some variable that has to leave the basis in exchange for a nonbasic one).

If the optimal solutions for each respective parameter interval are unique, then PSM and our approach necessarily produce the same solution path. However, the paths may differ if multiple optimal solutions ocuur, as the underlying algorithmic concepts are different: For one thing, we operate in the original variable space (n primal and m dual variables versus $2n + 2m$ variables in the above parameterized LP), and thus avoid doubling the dimensions. Moreover, in each iteration, PSM is restricted to moving to an adjacent basis and, in particular, can get stuck at a certain parameter value for several iterations (namely when several pivot steps are needed to eventually arrive at a new basis that allows to further reduce the parameter).

Regarding implementation, PSM is subject to all advantages and drawbacks that come with any simplex method, e.g., its basic version (as described in [44]) may cycle and hence not even terminate, special care needs to be taken to compute and maintain numerically stable basis matrix factorizations, etc. Our approach is straightforward to implement, but requires access to an LP solver for subproblem optimization—given the large selection of sophisticated LP solvers (both proprietary and freely available) to choose from, we actually consider this a feature, not a disadvantage. In particular, this allows us to use the active-set LP strategy described in Chapter 4 that turns out to be particularly well-suited to the subproblems occurring during our method. At least in case of multiple optimal solutions, both PSM and our homotopy method are naturally influenced by choices made for crucial steps (i.e., pivoting rules for PSM and LP subproblem solver choice in our implementation), which makes a direct numerical comparison somewhat meaningless. Hence, we do not delve into this subject further.

5.3 Extension to Non-Uniform Constraints

The fact that the ℓ_∞-norm constraint in (P_δ) has the representation (5.1) raises the question whether ℓ_1-HOUDINI can be modified to handle more general constraints of the form $\boldsymbol{\alpha} \leq \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{\beta}$, given that the associated problem has a feasible solution. In Subsection 5.3.1 (and previously in [5]), we show that it is not even necessary to modify ℓ_1-HOUDINI in order to solve the generalized problem. If we have $-\boldsymbol{\infty} < \boldsymbol{\alpha} < \boldsymbol{\beta} < \boldsymbol{\infty}$ (two-sided inequality constraints), it will turn out that there exist a diagonal matrix $\boldsymbol{G}$ as well as a right-hand side $\hat{\boldsymbol{b}}$ and a $\hat{\delta} > 0$ such that the modified constraint can be recast as $\|\boldsymbol{G}(\boldsymbol{A}\boldsymbol{x} - \hat{\boldsymbol{b}})\|_\infty \leq \hat{\delta}$, i.e., ℓ_1-HOUDINI can be applied to the related problem directly without adapting the algorithm. In Subsection 5.3.2, we go one step further and show that ℓ_1-HOUDINI can be adapted in order to handle ℓ_1-norm minimization problems with one-sided inequality constraints and equality constraints as well.

5.3.1 Two-Sided Inequality Constraints

We consider the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \boldsymbol{\alpha} \leq \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{\beta} \tag{5.8}$$

with $-\boldsymbol{\infty} < \boldsymbol{\alpha} < \boldsymbol{\beta} < \boldsymbol{\infty}$ and assume that the associated feasible set is non-empty. As it does not necessarily hold that $\boldsymbol{\alpha} = -\boldsymbol{\beta}$ and we explicitly allow both vectors to be non-constant (in the sense that $\boldsymbol{\alpha}, \boldsymbol{\beta} \neq c \cdot \mathbb{1}$), we refer to this type of constraints as *non-uniform constraints*. First, we make the observation that

$$\boldsymbol{\alpha} \leq \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{\beta} \quad \Leftrightarrow \quad \underbrace{\boldsymbol{\alpha} - \tfrac{\boldsymbol{\alpha}+\boldsymbol{\beta}}{2}}_{=:-\boldsymbol{\gamma}} \leq \boldsymbol{A}\boldsymbol{x} - (\underbrace{\boldsymbol{b} + \tfrac{\boldsymbol{\alpha}+\boldsymbol{\beta}}{2}}_{=:\hat{\boldsymbol{b}}}) \leq \underbrace{\boldsymbol{\beta} - \tfrac{\boldsymbol{\alpha}+\boldsymbol{\beta}}{2}}_{=\boldsymbol{\gamma}}. \tag{5.9}$$

By assumption, it holds that $\gamma_i \neq 0$ for all i. Hence, if we choose an arbitrary $\hat{\delta} > 0$ and set $\boldsymbol{G} \coloneqq \hat{\delta}\,\mathrm{Diag}(1/\gamma_1, \ldots, 1/\gamma_m)$, then scaling (5.9) by $\boldsymbol{G}$ results in the equivalent constraint

$$-\hat{\delta}\mathbb{1} \leq \boldsymbol{G}(\boldsymbol{A}\boldsymbol{x} - \hat{\boldsymbol{b}}) \leq \hat{\delta}\mathbb{1} \quad \Leftrightarrow \quad \|\boldsymbol{G}(\boldsymbol{A}\boldsymbol{x} - \hat{\boldsymbol{b}})\|_\infty \leq \hat{\delta}. \tag{5.10}$$

It follows that (5.8) can be rewritten in the form (P_δ) (and vice versa) and therefore, we can find a solution of (5.8) by applying ℓ_1-HOUDINI to the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \|\boldsymbol{G}(\boldsymbol{A}\boldsymbol{x} - \hat{\boldsymbol{b}})\|_\infty \leq \hat{\delta}. \tag{5.11}$$

5.3.2 Arbitrary Linear Constraints

Our final goal in this section is to adapt ℓ_1-HOUDINI to the problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \delta \\ & \boldsymbol{C}\boldsymbol{x} - \boldsymbol{d} = \boldsymbol{0} \\ & \boldsymbol{E}\boldsymbol{x} - \boldsymbol{f} \leq \boldsymbol{0}. \end{aligned} \tag{5.12}$$

Therein, we explicitly omit the type of constraints that we considered in (5.8) because we have just seen that these can be reformulated as ℓ_∞-norm constraints. Additionally, the problem (5.12) features equality constraints as well as one-sided inequality constraints corresponding to the case $\boldsymbol{\alpha} = -\boldsymbol{\infty}$ and $\boldsymbol{\beta} = \mathbf{0}$ in (5.8). Note that the reverse case with $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \boldsymbol{\infty}$ is as well captured by (5.12). To that end, we can simply switch the signs of all coefficients in the respective inequations.

In order to make our above homotopy principle applicable, we consider the problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \theta + \delta \\ & \|\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}\|_\infty \leq \theta \\ & \boldsymbol{E}\boldsymbol{x} - \boldsymbol{f} \leq \theta\mathbb{1} \end{aligned} \tag{5.13}$$

with some $\theta \geq 0$ which has, in case $\theta \geq \max\{\|\boldsymbol{b}\|_\infty - \delta, \|\boldsymbol{d}\|_\infty, \|\boldsymbol{f}^-\|_\infty\}$, the optimal solution $\boldsymbol{x}^\star = \mathbf{0}$. Provided that there exists a feasible point for (5.12), the well-known idea is to send the parameter θ in (5.13) to zero so as to obtain the associated solution path and, finally, an optimal solution of (5.12).

Corollary 46. *Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{C} \in \mathbb{R}^{p\times n}$, $\boldsymbol{d} \in \mathbb{R}^p$, $\boldsymbol{E} \in \mathbb{R}^{q\times n}$, $\boldsymbol{f} \in \mathbb{R}^q$ and $\delta, \theta \geq 0$ such that the problem (5.12) is feasible. Then, it holds that*

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_\infty \leq \theta + \delta \\ & \|\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}\|_\infty \leq \theta \\ & \boldsymbol{E}\boldsymbol{x} - \boldsymbol{f} \leq \theta\mathbb{1} \end{aligned} \tag{P_θ}$$

$$\begin{aligned} = \max_{(\boldsymbol{y},\boldsymbol{z},\boldsymbol{u})\in\mathbb{R}^{m\times k\times \ell}} \quad & (\delta+\theta)\|\boldsymbol{y}\|_1 + \theta\|\boldsymbol{z}\|_1 + \theta\mathbb{1}^\top\boldsymbol{u} + \boldsymbol{b}^\top\boldsymbol{y} + \boldsymbol{d}^\top\boldsymbol{z} + \boldsymbol{f}^\top\boldsymbol{u} \\ \text{s.t.} \quad & \|\boldsymbol{A}^\top\boldsymbol{y} + \boldsymbol{C}^\top\boldsymbol{z} + \boldsymbol{E}^\top\boldsymbol{u}\|_\infty \leq 1 \\ & \boldsymbol{u} \geq \mathbf{0} \end{aligned} \tag{D_θ}$$

and the minimum and the maximum are attained at $\boldsymbol{x}^\star$ and $(\boldsymbol{y}^\star, \boldsymbol{z}^\star, \boldsymbol{u}^\star)$ if and only if

$$-\boldsymbol{A}^\top\boldsymbol{y}^\star - \boldsymbol{C}^\top\boldsymbol{z}^\star - \boldsymbol{E}^\top\boldsymbol{u}^\star \in \partial\|\boldsymbol{x}^\star\|_1 \quad \textit{and} \quad \begin{aligned} \boldsymbol{A}\boldsymbol{x}^\star - \boldsymbol{b} &\in (\theta+\delta)\partial\|\boldsymbol{y}^\star\|_1 \\ \boldsymbol{C}\boldsymbol{x}^\star - \boldsymbol{d} &\in \theta\partial\|\boldsymbol{z}^\star\|_1 \\ \boldsymbol{E}\boldsymbol{x}^\star - \boldsymbol{f} &\leq \theta\mathbb{1} \\ (\boldsymbol{E}\boldsymbol{x}^\star - \boldsymbol{f} - \theta\mathbb{1}) \odot \boldsymbol{u}^\star &= \mathbf{0} \\ \boldsymbol{u}^\star &\geq \mathbf{0} \end{aligned} \quad . \tag{5.14}$$

Proof. Corollary 46 generalizes Theorem 12 to problems with equality constraints and one-sided inequality constraints. The argumentation is completely analogous to the proof of Theorem 12 (and preceding statements in Section 2.2). In particular, both (P_θ) and (D_θ) can be recast as linear programs which are dual to each other. The main difference is that we must apply Theorems 2 and 5, respectively, with the linear transformation $\boldsymbol{K} = [\boldsymbol{A}^\top\ \boldsymbol{C}^\top\ \boldsymbol{E}^\top]^\top$ and the function

$$g(\boldsymbol{K}\boldsymbol{x}) = g(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{C}\boldsymbol{x}, \boldsymbol{E}\boldsymbol{x}) = I_{\|\cdot-\boldsymbol{b}\|_\infty\leq\theta+\delta}(\boldsymbol{A}\boldsymbol{x}) + I_{\|\cdot-\boldsymbol{d}\|_\infty\leq\theta}(\boldsymbol{C}\boldsymbol{x}) + I_{(\cdot)-\boldsymbol{f}\leq\theta\mathbb{1}}(\boldsymbol{E}\boldsymbol{x}) \tag{5.15}$$

which has the fenchel conjugate

$$g^*(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}) = (\theta + \delta)\|\boldsymbol{y}\|_1 + \boldsymbol{b}^\top \boldsymbol{y} + \theta\|\boldsymbol{z}\|_1 + \boldsymbol{d}^\top \boldsymbol{z} + (\theta \mathbb{1} + \boldsymbol{f})^\top \boldsymbol{u} + I_{(\cdot)\geq \mathbf{0}}(\boldsymbol{u}). \tag{5.16}$$

The conditions on the right side of (5.14) (in particular, the three conditions on the lower end) are finally true because

$$\partial I_{(\cdot)\geq 0}(\boldsymbol{u})_k = \begin{cases} \mathbb{R}_-, & u_k = 0 \\ \{0\}, & u_k > 0 \\ \emptyset, & u_k < 0. \end{cases} \tag{5.17}$$

□

Compared to (P_δ), the number of primal variables and the size of the ℓ_∞-norm constraint in the dual problem (both n) are identical. Therefore, we still refer to the primal support as S and to the dual active set as Σ in the following. However, we have additional constraints in (P_θ) and additional dual variables in (D_θ). Consequently, we define the sets

$$\begin{aligned} W_{\boldsymbol{y}} &= \{i : |\boldsymbol{a}_i^\top \boldsymbol{x} - b_i| = \theta + \delta\} & \Omega_{\boldsymbol{y}} &= \{i : y_i \neq 0\} \\ W_{\boldsymbol{z}} &= \{r : |\boldsymbol{c}_r^\top \boldsymbol{x} - d_r| = \theta\} & \Omega_{\boldsymbol{z}} &= \{k : z_r \neq 0\} \\ W_{\boldsymbol{u}} &= \{s : \boldsymbol{e}_s^\top \boldsymbol{x} - f_s = \theta\} & \Omega_{\boldsymbol{u}} &= \{s : u_s > 0\}. \end{aligned} \tag{5.18}$$

Nevertheless, to keep our notation on a reasonably low level, we omit the subscripts in the following, i.e., $\boldsymbol{y}_W$ refers to $\boldsymbol{y}_{W_{\boldsymbol{y}}}$, $\boldsymbol{C}^\Omega$ refers to $\boldsymbol{C}^{\Omega_{\boldsymbol{z}}}$, and so on.

Analogous to Section 3.2, where we derived our homotopy approach for (P_δ) starting from the primal-dual optimality conditions (3.1), we obtain a generalized homotopy approach for (P_θ) if we proceed from the optimality conditions (5.14). We state the following two corollaries without proofs as they are straightforward extensions of Lemmas 13 and 14. Note that we reuse the labels (C_D^k) and (C_P^k) from Section 3.2 to denote the generalized conditions in the current section.

Corollary 47. *Let $\boldsymbol{x}^k$ be an optimal solution of (P_{θ^k}). Then $\boldsymbol{x}^k$ and $(\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{u}^{k+1})$ form an optimal quadruplet for (P_{θ^k}) if and only if the latter satisfy*

$$\begin{aligned} -\boldsymbol{A}_{S_k}^\top \boldsymbol{y} - \boldsymbol{C}_{S_k}^\top \boldsymbol{z} - \boldsymbol{E}_{S_k}^\top \boldsymbol{u} &= \operatorname{sign}(\boldsymbol{x}_{S_k}^k) \\ -\mathbb{1} \leq -\boldsymbol{A}_{S_k^c}^\top \boldsymbol{y} - \boldsymbol{C}_{S_k^c}^\top \boldsymbol{z} - \boldsymbol{E}_{S_k^c}^\top \boldsymbol{u} &\leq \mathbb{1} \\ -\operatorname{sign}(\boldsymbol{A}^{W_k} \boldsymbol{x}^k - \boldsymbol{b}_{W_k}) \odot \boldsymbol{y}_{W_k} &\leq \mathbf{0} \\ -\operatorname{sign}(\boldsymbol{C}^{W_k} \boldsymbol{x}^k - \boldsymbol{d}_{W_k}) \odot \boldsymbol{z}_{W_k} &\leq \mathbf{0} \\ -\mathbb{1} \odot \boldsymbol{u}_{W_k} &\leq \mathbf{0} \\ \boldsymbol{y}_{W_k^c}, \boldsymbol{z}_{W_k^c}, \boldsymbol{u}_{W_k^c} &= \mathbf{0}. \end{aligned} \tag{C_D^k}$$

Corollary 48. *Let $(\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{u}^{k+1})$ be an optimal solution of (D_{θ^k}). Then $\boldsymbol{x}^{k+1}$ and $(\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{u}^{k+1})$ form an optimal quadruplet for ($P_{\theta^k - t^{k+1}}$) with $\theta^k - t^{k+1} \geq 0$ if and*

only if $\boldsymbol{x}^{k+1}$ *and* t^{k+1} *form a solution of*

$$\begin{aligned}
\boldsymbol{A}^{\Omega_{k+1}}\boldsymbol{x} - \boldsymbol{b}_{\Omega_{k+1}} &= (\theta^k + \delta - t)\operatorname{sign}(\boldsymbol{y}^{k+1}_{\Omega_{k+1}}) \\
\boldsymbol{C}^{\Omega_{k+1}}\boldsymbol{x} - \boldsymbol{d}_{\Omega_{k+1}} &= (\theta^k - t)\operatorname{sign}(\boldsymbol{z}^{k+1}_{\Omega_{k+1}}) \\
\boldsymbol{E}^{\Omega_{k+1}}\boldsymbol{x} - \boldsymbol{f}_{\Omega_{k+1}} &= (\theta^k - t)\mathbb{1} \\
-(\theta^k + \delta - t)\mathbb{1} \leq \boldsymbol{A}^{\Omega^c_{k+1}}\boldsymbol{x} - \boldsymbol{b}_{\Omega^c_{k+1}} &\leq (\theta^k + \delta - t)\mathbb{1} \\
-(\theta^k - t)\mathbb{1} \leq \boldsymbol{C}^{\Omega^c_{k+1}}\boldsymbol{x} - \boldsymbol{d}_{\Omega^c_{k+1}} &\leq (\theta^k - t)\mathbb{1} \qquad (\mathrm{C}^k_P) \\
\boldsymbol{E}^{\Omega^c_{k+1}}\boldsymbol{x} - \boldsymbol{f}_{\Omega^c_{k+1}} &\leq (\theta^k - t)\mathbb{1} \\
(\boldsymbol{A}^\top_{\Sigma_{k+1}}\boldsymbol{y}^{k+1} + \boldsymbol{C}^\top_{\Sigma_{k+1}}\boldsymbol{z}^{k+1} + \boldsymbol{E}^\top_{\Sigma_{k+1}}\boldsymbol{u}^{k+1}) \odot \boldsymbol{x}_{\Sigma_{k+1}} &\leq \boldsymbol{0} \\
\boldsymbol{x}_{\Sigma^c_{k+1}} &= \boldsymbol{0} \\
t &\leq \theta^k.
\end{aligned}$$

Recall that, in terms of our original homotopy method, the conditions (C^k_D) and (C^k_P) serve as constraints in the dual and primal update problem, respectively. Hence, at this point, the missing building blocks to complete the adapted version of ℓ_1-HOUDINI are the objective functions for the dual and primal update problems. It turns out that we can essentially use the same objective functions as before, except that we have to add appropriate terms corresponding to $\boldsymbol{z}$ and $\boldsymbol{u}$ in case of the dual update. Accordingly, we choose

$$(\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{u}^{k+1}) \in \operatorname*{arg\,min}_{(\boldsymbol{y},\boldsymbol{z},\boldsymbol{u})\in\mathbb{R}^{m+p+q}} \boldsymbol{\psi}_{\boldsymbol{y}}^\top\boldsymbol{y} + \boldsymbol{\psi}_{\boldsymbol{z}}^\top\boldsymbol{z} + \boldsymbol{\psi}_{\boldsymbol{u}}^\top\boldsymbol{u} \quad \text{s.t. } (\boldsymbol{y},\boldsymbol{z},\boldsymbol{u}) \text{ satisfy } (\mathrm{C}^k_D),\ (\mathrm{U}^k_D)$$

where $\boldsymbol{\psi}_{\boldsymbol{y}} = -\operatorname{sign}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$, $\boldsymbol{\psi}_{\boldsymbol{z}} = -\operatorname{sign}(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d})$ and $\boldsymbol{\psi}_{\boldsymbol{u}} = -\mathbb{1}$. In case of the primal update, we search for

$$(\boldsymbol{x}^{k+1}, t^{k+1}) \in \operatorname*{arg\,max}_{(\boldsymbol{x},t)\in\mathbb{R}^{n+1}} t \quad \text{s.t. } (\boldsymbol{x}, t) \text{ satisfy } (\mathrm{C}^k_P). \qquad (\mathrm{U}^k_P)$$

The entire iterative scheme is illustrated in Algorithm 3. Concerning finite termination, note that the statements from Subsections 3.3 and 3.4 have equivalents in terms of the generalized problems (P_θ) and (D_θ) and Algorithm 3, respectively. Hence, Algorithm 3 returns an optimal point after a finite number of iterations. Since the structure of the dual and primal update problems is essentially the same as before, it is possible to derive an active-set method analogous to Sections 4.2–4.4 as well.

Finally, let us remark that the relationship between (P_δ) and linear programming extends, in a sense, both ways: On the one hand, we have shown that (P_δ) is equivalent to the linear program (5.2) and can consequently be solved directly using an arbitrary LP solver. On the other hand, suppose that a linear program

$$\begin{aligned}
\min_{\boldsymbol{x}\in\mathbb{R}^n}\ \boldsymbol{c}^\top\boldsymbol{x} \quad \text{s.t. } \boldsymbol{C}\boldsymbol{x} - \boldsymbol{d} &= \boldsymbol{0} \\
\boldsymbol{E}\boldsymbol{x} - \boldsymbol{f} &\leq \boldsymbol{0} \\
\boldsymbol{x} &\geq \boldsymbol{0}
\end{aligned} \qquad (5.19)$$

with $\boldsymbol{c} > \mathbf{0}$ is given. After the substitution $\tilde{\boldsymbol{x}} \coloneqq \operatorname{Diag}(\boldsymbol{c})\boldsymbol{x}$, we see that (5.19) is equivalent to the problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^n} \ \mathbb{1}^\top \tilde{\boldsymbol{x}} \quad \text{s.t. } & \boldsymbol{C} \operatorname{Diag}(\boldsymbol{c})^{-1} \tilde{\boldsymbol{x}} - \boldsymbol{d} = \mathbf{0} \\ & \boldsymbol{E} \operatorname{Diag}(\boldsymbol{c})^{-1} \tilde{\boldsymbol{x}} - \boldsymbol{f} \leq \mathbf{0} \\ & \tilde{\boldsymbol{x}} \geq \mathbf{0} \end{aligned} \tag{5.20}$$

and hence, also to the problem

$$\begin{aligned} \min_{\tilde{\boldsymbol{x}} \in \mathbb{R}^n} \ \|\tilde{\boldsymbol{x}}\|_1 \quad \text{s.t.} \quad & \boldsymbol{C} \operatorname{Diag}(\boldsymbol{c})^{-1} \tilde{\boldsymbol{x}} - \boldsymbol{d} = \mathbf{0} \\ & \begin{bmatrix} \boldsymbol{E} \operatorname{Diag}(\boldsymbol{c})^{-1} \\ -\boldsymbol{I}_n \end{bmatrix} \tilde{\boldsymbol{x}} - \begin{pmatrix} \boldsymbol{f} \\ \mathbf{0} \end{pmatrix} \leq \mathbf{0} \end{aligned} \tag{5.21}$$

which has again the form of (5.12). Thus, after an appropriate rescaling of the rows of $\boldsymbol{C}$ and $\boldsymbol{E}$, our generalized version of ℓ_1-HOUDINI can be applied to any linear program with strictly positive objective function coefficients.

Input: $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{C} \in \mathbb{R}^{p\times n}$, $\boldsymbol{d} \in \mathbb{R}^p$, $\boldsymbol{E} \in \mathbb{R}^{q\times n}$, $\boldsymbol{f} \in \mathbb{R}^q$, $\delta > 0$
Output: solution $\boldsymbol{x}^\star$ to problem (5.12)

// Initialization:
$\theta^0 \leftarrow \max\{\|\boldsymbol{b}\|_\infty - \delta, \|\boldsymbol{d}\|_\infty, \|\boldsymbol{f}^-\|_\infty\}$
$\boldsymbol{x}^0 \leftarrow \boldsymbol{0}$
$S_0 \leftarrow \emptyset$
$W_{\boldsymbol{y},0} \leftarrow \{i : |b_i| = \theta^0 + \delta\}$
$W_{\boldsymbol{z},0} \leftarrow \{r : |d_r| = \theta^0\}$
$W_{\boldsymbol{u},0} \leftarrow \{s : f_s = -\theta^0\}$
$k \leftarrow 0$
repeat
  // Dual update:
  $(\boldsymbol{y}^{k+1}, \boldsymbol{z}^{k+1}, \boldsymbol{u}^{k+1}) \leftarrow$ solution of (U_D^k)
  $\Omega_{\boldsymbol{y},k+1} \leftarrow \left\{i : y_i^{k+1} \neq 0\right\}$
  $\Omega_{\boldsymbol{z},k+1} \leftarrow \left\{r : z_r^{k+1} \neq 0\right\}$
  $\Omega_{\boldsymbol{u},k+1} \leftarrow \left\{s : u_s^{k+1} > 0\right\}$
  $\Sigma_{k+1} \leftarrow \left\{j : |\boldsymbol{A}_j^\top \boldsymbol{y}^{k+1} + \boldsymbol{C}_j^\top \boldsymbol{z}^{k+1} + \boldsymbol{E}_j^\top \boldsymbol{u}^{k+1}| = 1\right\}$
  // Primal update:
  $(\boldsymbol{x}^{k+1}, t^{k+1}) \leftarrow$ solution of (C_P^k)
  $\theta^{k+1} \leftarrow \theta^k - t^{k+1}$
  $S_{k+1} \leftarrow \left\{j : x_j^{k+1} \neq 0\right\}$
  $W_{\boldsymbol{y},k+1} \leftarrow \left\{i : |\boldsymbol{a}_i^\top \boldsymbol{x}^{k+1} - b_i| = \theta^{k+1} + \delta\right\}$
  $W_{\boldsymbol{z},k+1} \leftarrow \left\{r : |\boldsymbol{c}_r^\top \boldsymbol{x}^{k+1} - d_r| = \theta^{k+1}\right\}$
  $W_{\boldsymbol{u},k+1} \leftarrow \left\{s : \boldsymbol{e}_s^\top \boldsymbol{x}^{k+1} - f_i = \theta^{k+1}\right\}$
  $k \leftarrow k + 1$
until $\theta^k = 0$
return $\boldsymbol{x}^\star = \boldsymbol{x}^k$

Algorithm 3: Variant of ℓ_1-HOUDINI adapted to problem (5.12).

6 Applications

6.1 Speech Coding

In this section, we pick up on an application from the field of *speech coding* that was originally addressed in [4]. Speech coding describes how analog speech signals can efficiently be represented in the digital domain, for instance for storage and transmission. Figure 6.1 illustrates a typical speech coding scheme capturing the example of mobile telephony. At the transmitting end, the sender speaks into the microphone of a cell phone. There, the analog speech signal is converted to a digital signal and afterwards transmitted to the mobile phone of the sender where it is finally decoded and made audible to the receiver. Taken together, the encoder and the decoder are usually referred to as a *speech codec*.

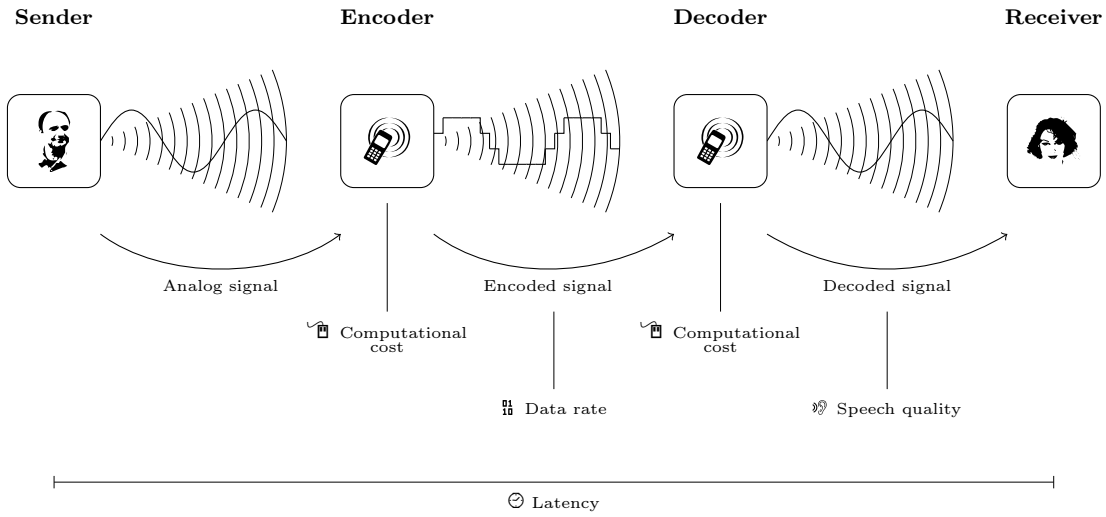

Figure 6.1: Speech coding using the example of mobile telephony.

The design of a speech codec needs to be fit to the application at hand because different desirable properties are potentially conflicting. For example, sophisticated algorithms in the encoding and decoding steps can of course help to yield a good speech quality at the receiving end. However, this may go along with relatively high computational cost for encoding and decoding as well as an inacceptable latency. Moreover, a high data rate, which is tendencially also beneficial for speech quality, is associated with a relatively high energy demand at the sending device. In case the sending device is battery driven, this calls for a preferably low data rate. In view of these aspects, the

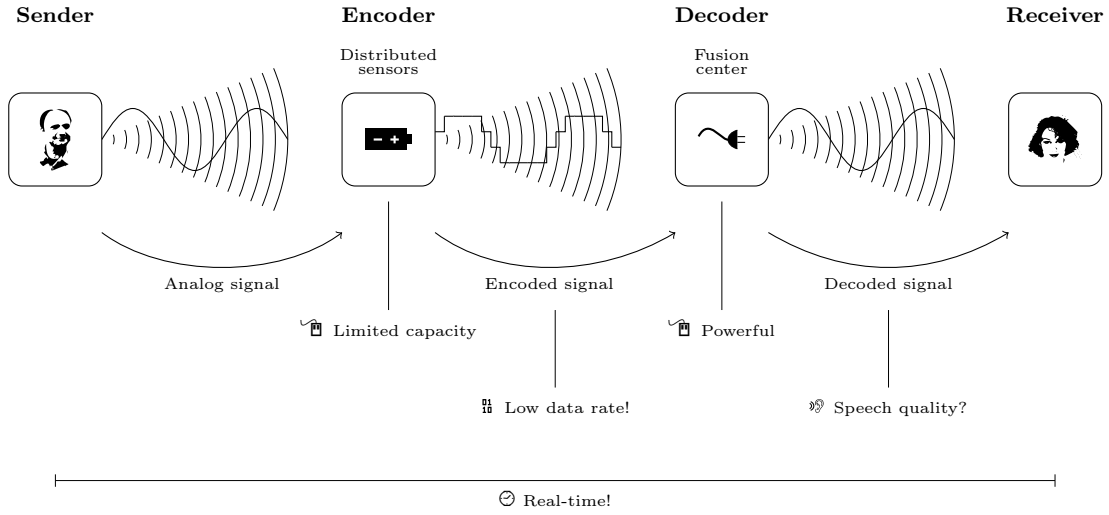

Figure 6.2: Speech coding using the example of wireless acoustic sensor networks.

goal of research in speech coding is to find algorithms that give the best possible trade-off between computational complexity for encoding and decoding, the required data rate, algorithmic latency and speech quality (see [45]).

In this section, we focus on the example of *wireless acoustic sensor networks* which is illustrated in Figure 6.2. The idea of wireless acoustic sensor networks is that many cheap, small and potentially battery driven acoustic sensors are spread through a room or even a house of interest. A speech signal that is captured by a sensor is encoded and afterwards sent to a central processor called the *fusion center* where the decoding is done. As opposed to the sensors, the fusion center is equipped with a plug and a powerful processor. Hence, the resources between the sender and the receiver are quite unbalanced which imposes the following requirements on a related speech coding scheme: The computational effort for the encoding procedure as well as the data rate for transmission need to be small because the sensors have relatively low computational and battery capacities. In contrast, the process of decoding at the fusion center can be much more complex with the restriction that the whole coding and decoding procedure can be accomplished in real-time, i.e., the latency must be very small. Now, the objective is to design a speech codec that meets all of these requirements and that provides a high speech quality at the receiving end. In the following, we describe the algorithmic approach that was proposed in [4] and show that it can be allocated to the class of *statistical-model-based algorithms* (see [27]).

6.1.1 Encoding

We model an analog speech signal as a function $f : T \to (-1, 1)$, where $T \subseteq \mathbb{R}$ is some time domain. In the first step, f is sampled at equidistant points $\{t_1, \ldots, t_N\} \subseteq T$ resulting in the sampled signal $\boldsymbol{f} \in (-1, 1)^N$ with $f_j \coloneqq f(t_j)$ for $j = 1, \ldots, N$ (see

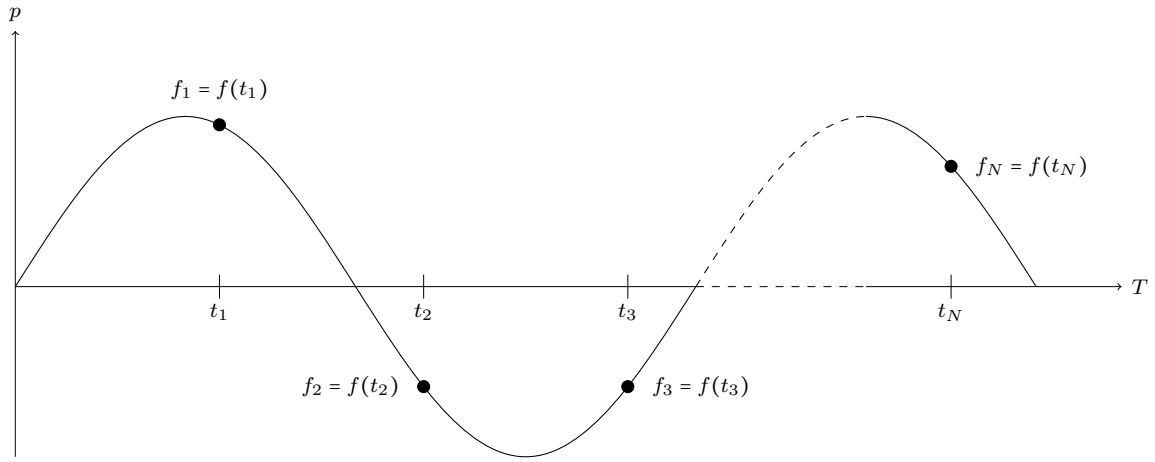

Figure 6.3: Sampling of an analog signal.

Figure 6.3).

In a second step, the signal $\boldsymbol{f}$ is quantized with some word length $w \in \mathbb{N}$ which corresponds to the number of bits per sample that is available for transmission. To that end, the interval $[0, 1)$ is subdivided into 2^{w-1} pairwise disjoint intervals. Hence, it holds that $[0, 1) = I_1 \cup I_2 \cup \cdots \cup I_{2^{w-1}}$ (w.l.o.g. we further assume that $I_1 \leq I_2 \leq \cdots \leq I_{2^{w-1}}$ which implies $0 \in I_1$). For each interval, we choose an associated quantization level $0 < \Delta_l \in I_l$. Therewith, the quantization function $Q : (-1, 1)^n \to (-1, 1)^n$ is defined componentwise as

$$Q(\boldsymbol{f})_j \coloneqq \operatorname{sign}^+(f_j)\Delta_l \quad \text{if} \quad |f_j| \in I_l. \tag{6.1}$$

Note that Q is a so called *mid-riser* which refers to the fact that each zero component of $\boldsymbol{f}$ is quantized to the value $\Delta_1 > 0$. From the above definition, it follows that Q is odd in each component (at least if we neglect that $Q(0) \neq 0$) and that it maps the interval $(-1, 1)$ to 2^w different quantization levels (see Figure 6.4).

6.1.2 Decoding

In the decoding step, we are faced with the situation that we are aware of the quantized signal $Q(\boldsymbol{f})$ but not of $\boldsymbol{f}$ itself. Without any further knowledge, we can not expect to recover the original signal exactly because there exist infinitely many signals $\hat{\boldsymbol{f}}$ satisfying

$$Q(\hat{\boldsymbol{f}}) = Q(\boldsymbol{f}). \tag{6.2}$$

However, the assumption that each (in some sense) good approximation of the original signal satisfies (6.2) seems to be reasonable. In order to make the search for a reconstruction feasible, we further assume that there exist an a priori known dictionary $\boldsymbol{D}$ as well as a sparse coefficient vector $\boldsymbol{a}$ such that

$$\boldsymbol{f} = \boldsymbol{D}\boldsymbol{a}. \tag{6.3}$$

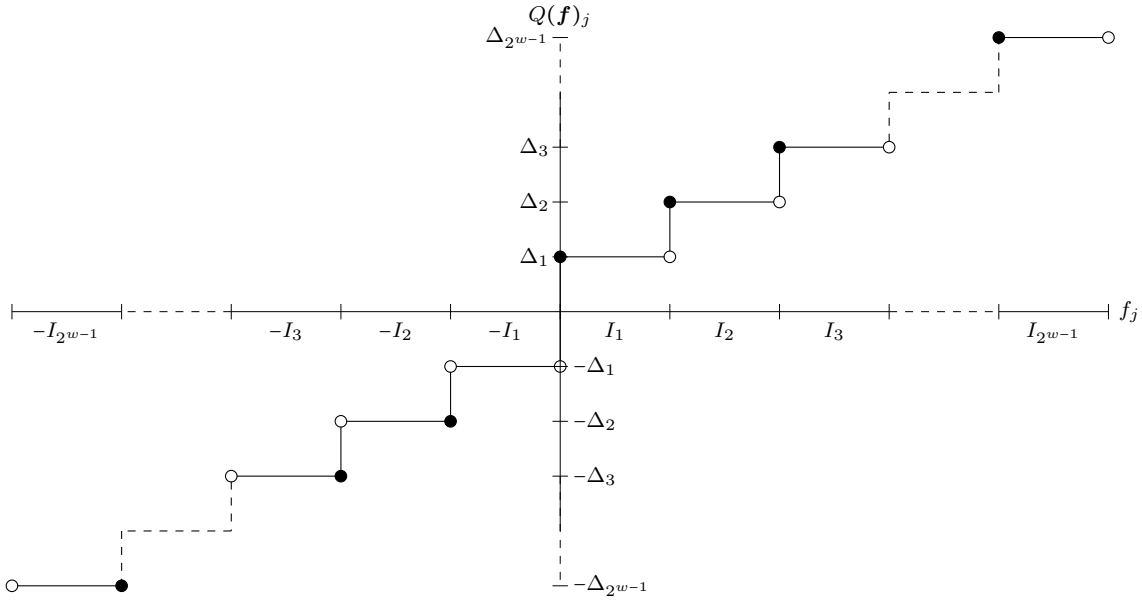

Figure 6.4: Example of a quantization function.

As the ℓ_1-norm used as an objective function is known to prefer sparse solutions of linear equation systems (see [17]), the sparsity assumption on $\boldsymbol{a}$ together with (6.2) and (6.3) leads us to the optimization problem

$$\inf_{\boldsymbol{a}\in\mathbb{R}^n} \|\boldsymbol{a}\|_1 \quad \text{s.t. } Q(\boldsymbol{D}\boldsymbol{a}) = Q(\boldsymbol{f}). \tag{6.4}$$

Therein, we take the infimum instead of the minimum because the set of points satisfying the constraint is not necessarily closed. Therefore, the infimum might not be attained. However, the convex set

$$Q^{-1}(\boldsymbol{q}) \coloneqq \operatorname{cl}\left(\{\boldsymbol{\phi} \in (-1,1)^n \mid \forall j \in \{1,\dots,n\} : |q_j| = \Delta_l \Rightarrow \phi_j \in \operatorname{sign}(q_j) I_l\}\right) \tag{6.5}$$

with $\boldsymbol{q} \coloneqq Q(\boldsymbol{f})$ is equal to the closure of the feasible set in (6.4). Hence, the assignment

$$\hat{\boldsymbol{a}} \in \operatorname*{arg\,min}_{\boldsymbol{a}\in\mathbb{R}^n} \|\boldsymbol{a}\|_1 \quad \text{s.t. } \boldsymbol{D}\boldsymbol{a} \in Q^{-1}(\boldsymbol{q}) \tag{6.6}$$

can be considered a relaxation of (6.4) and is in particular well-defined. Finally, we choose $\hat{\boldsymbol{f}} \coloneqq \boldsymbol{D}\hat{\boldsymbol{a}}$ as our approximation of the original signal. Although the function Q is non-linear, it follows from (6.5) that the constraint in (6.6) actually decomposes into a set of linear inequalities. In the following, we consider two different types of quantization functions which are *uniform* and *non-uniform* quantization functions. If the dictionary $\boldsymbol{D}$ is given, it only remains to determine the sets $Q^{-1}(\boldsymbol{q})$ subject to the respective quantization functions, before we can apply (6.6).

6.1.3 Uniform Quantization

In the first place, we consider uniform quantization functions where the interval $(-1,1)$ is subdivided into 2^w intervals of equal length $\Delta \coloneqq 2^{-w+1}$. The respective intervals

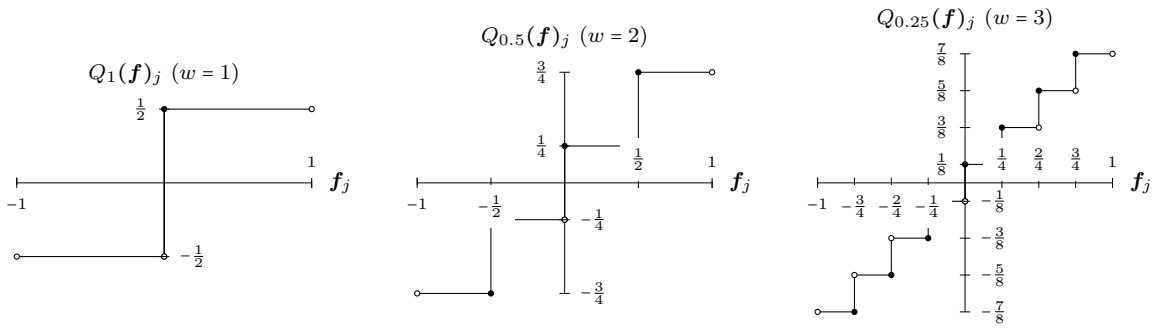

Figure 6.5: Uniform quantization

on the positive axis are $I_l \coloneqq [(l-1)\Delta, l\Delta)$ for $l = 1, \ldots, 2^{w-1}$ and the corresponding quantization levels are the center points $\Delta_l \coloneqq (l - \frac{1}{2})\Delta$. This kind of quantization function has a componentwise closed-form representation which is

$$Q(\boldsymbol{f})_j = Q_\Delta(\boldsymbol{f})_j \coloneqq \operatorname{sign}^+(f_j)\Delta\left(\left\lfloor \frac{|f_j|}{\Delta} \right\rfloor + \frac{1}{2}\right) \tag{6.7}$$

(see Figure 6.5). As the intervals can as well be written in the form $I_l = [\Delta_l - \frac{1}{2}, \Delta_l + \frac{1}{2})$, it follows that

$$Q_\Delta^{-1}(\boldsymbol{q}) = \left\{ \boldsymbol{\phi} \in (-1,1)^n \;\middle|\; \|\boldsymbol{\phi} - \boldsymbol{q}\|_\infty \leq \frac{\Delta}{2} \right\}. \tag{6.8}$$

Therewith, (6.6) becomes

$$\hat{\boldsymbol{a}} \;\in\; \operatorname*{arg\,min}_{\boldsymbol{a}\in\mathbb{R}^n} \|\boldsymbol{a}\|_1 \quad \text{s.t. } \|\boldsymbol{D}\boldsymbol{a} - \boldsymbol{q}\|_\infty \leq \frac{\Delta}{2} \tag{6.9}$$

which has exactly the form (P_δ). Hence, we can use ℓ_1-Houdini in order to find $\hat{\boldsymbol{a}}$ and afterwards determine the reconstructed signal as $\hat{\boldsymbol{f}} = \boldsymbol{D}\boldsymbol{a}$.

6.1.4 Non-uniform quantization

In the case of uniform quantization, the introduced *quantization noise* $\boldsymbol{\varepsilon} \coloneqq \boldsymbol{q} - \boldsymbol{f}$ is bounded above by $\|\boldsymbol{\varepsilon}\|_\infty \leq \frac{\Delta}{2}$. This is due to the fact that the quantization levels Δ_l are the centers of the respective intervals I_l which have uniform length Δ. In particular, the bound on ε_j does not depend on the concrete value of q_j, i.e., the expectable error does not depend on the amplitude of the signal.

However, according to experiments (see [41]), there is more information contained in the lower amplitudes of speech signals than in the higher amplitudes. This observation gives rise to the idea of quantizing the higher amplitudes of a signal more coarsely while the lower and more important amplitudes are subject to a finer quantization. In terms of the quantization intervals, this approach corresponds to increasing interval lengths $|I_1| < |I_2| < \cdots < |I_{2^{w-1}}|$.

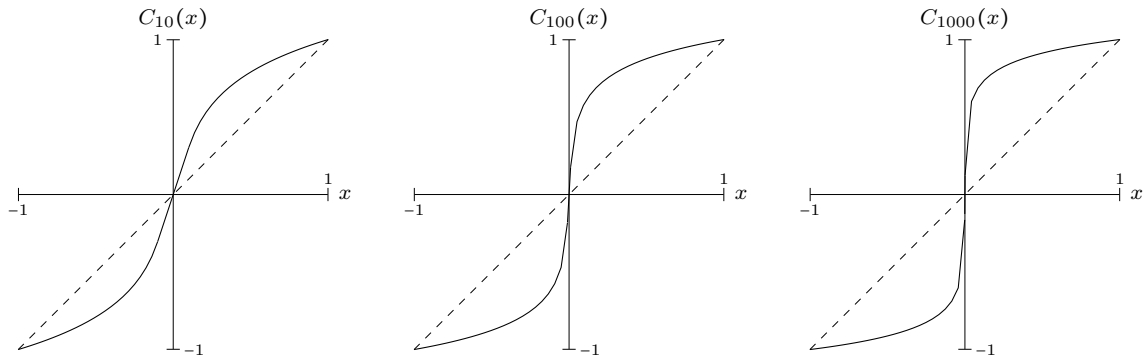

Figure 6.6: A-law compression function for $A \in \{10, 100, 1000\}$.

One way to introduce this kind of structure to the quantization intervals is to apply a *compression function* prior to a uniform quantization function. For some fixed $A \geq 1$, the *A-law* compression function $C_A : [-1, 1] \to [-1, 1]$, defined by

$$C_A(x) \coloneqq \begin{cases} \operatorname{sign}(x) \frac{1+\ln(A|x|)}{1+\ln(A)}, & 1 \geq |x| \geq \frac{1}{A}, \\ \operatorname{sign}(x) \frac{A|x|}{1+\ln(A)}, & \frac{1}{A} > |x| \geq 0, \end{cases} \tag{6.10}$$

is a common choice (see [39]). This function is odd, continuous and strictly monotonically increasing. It follows that C_A is invertible with

$$C_A^{-1}(y) = \begin{cases} A^{-1} \operatorname{sign}(y) \mathrm{e}^{|y|(1+\ln(A))-1}, & 1 \geq |y| \geq \frac{1}{1+\ln(a)}, \\ A^{-1}(1 + \ln(A))y, & \frac{1}{1+\ln(a)} > |y| \geq 0. \end{cases} \tag{6.11}$$

The examples in Figure 6.6 illustrate that the degree of compression increases with increasing values of A.

For ease of notation, we define $C_A(\boldsymbol{f})_j \coloneqq C_A(f_j)$ componentwise and therewith a non-uniform quantization function via

$$Q(\boldsymbol{f}) = Q_{\Delta,A}(\boldsymbol{f}) \coloneqq Q_\Delta(C_A(\boldsymbol{f})) \tag{6.12}$$

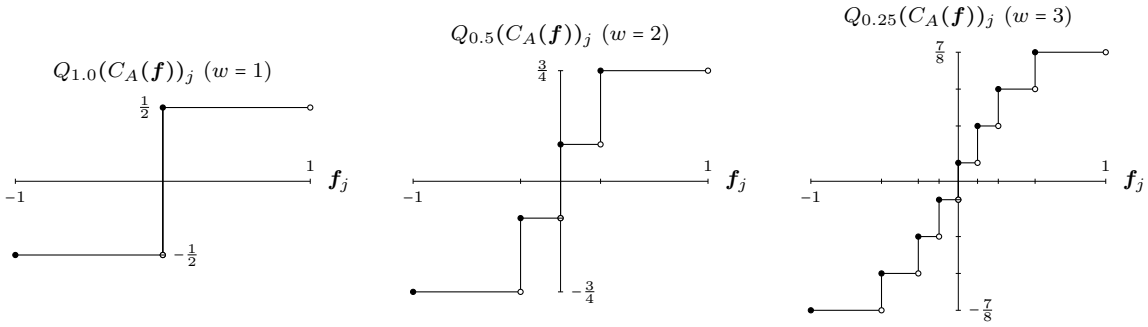

Figure 6.7: Non-uniform quantization functions with $A = 87.6$.

(see Figure 6.7 and note that $\Delta = 2^{-w+1}$ still depends on the word length w). To figure out $Q_{\Delta,A}^{-1}$ according to (6.5), we first observe that

$$Q_\Delta(C_A(\boldsymbol{f})) = \boldsymbol{q} \;\Rightarrow\; C_A(\boldsymbol{f}) \in Q_\Delta^{-1}(\boldsymbol{q}) \;\Leftrightarrow\; \boldsymbol{f} \in C_A^{-1}(Q_\Delta^{-1}(\boldsymbol{q})). \tag{6.13}$$

Using (6.8) and the monotonicity of C_A, we obtain that

$$Q_{\Delta,A}^{-1}(\boldsymbol{q}) = \left\{ \boldsymbol{\phi} \in (-1,1)^N \;\middle|\; C_A^{-1}\left(\boldsymbol{q} - \frac{\Delta}{2}\mathbb{1}\right) \leq \boldsymbol{\phi} \leq C_A^{-1}\left(\boldsymbol{q} + \frac{\Delta}{2}\mathbb{1}\right) \right\}. \tag{6.14}$$

Furter defining $\boldsymbol{\alpha} \coloneqq C_A^{-1}\left(\boldsymbol{q} - \frac{\Delta}{2}\mathbb{1}\right)$ and $\boldsymbol{\beta} \coloneqq C_A^{-1}\left(\boldsymbol{q} + \frac{\Delta}{2}\mathbb{1}\right)$, problem (6.6) turns into

$$\hat{\boldsymbol{a}} \;\in\; \operatorname*{arg\,min}_{\boldsymbol{a}\in\mathbb{R}^n} \|\boldsymbol{a}\|_1 \quad \text{s.t. } \boldsymbol{\alpha} \leq \boldsymbol{D}\boldsymbol{a} \leq \boldsymbol{\beta}. \tag{6.15}$$

Using the technique discussed in Subsection 5.3.1, with $\boldsymbol{b} = \boldsymbol{0}$, we can rewrite (6.15) as

$$\hat{\boldsymbol{a}} \;\in\; \operatorname*{arg\,min}_{\boldsymbol{a}\in\mathbb{R}^n} \|\boldsymbol{a}\|_1 \quad \text{s.t. } \| \operatorname{Diag}(\boldsymbol{\beta} - \boldsymbol{\alpha})^{-1} (2\boldsymbol{D}\boldsymbol{a} - (\boldsymbol{\beta} + \boldsymbol{\alpha})) \|_\infty \leq 1. \tag{6.16}$$

6.1.5 Numerical Experiments

Extensive numerical experiments investigating the impact of the estimating procedures (6.9) and (6.15) on the speech quality of decoded signals were already performed by the authors of [4]. To that end, both approaches were applied to 720 sentences from the IEEE corpus provided in [27] consisting of male speech and sampled at 16 kHz. The approach in [4] is to solve the problems (6.9) and (6.15) by using Chambolle-Pock's primal-dual method [12] with a fixed number of iterations and starting points $Q_\Delta(\boldsymbol{f})$ and $C_A^{-1}(Q_{\Delta,A}(\boldsymbol{f}))$, respectively.

The experimental setting in [4] is as follows: As a first step, the speech signal $\boldsymbol{f}$ is quantized using a uniform or non-uniform quantization function as described above. Then, the quantized signal is split into overlapping sub-signals $\boldsymbol{q}^t$ to which Chambolle-Pock's algorithm is applied, yielding coefficient vectors $\hat{\boldsymbol{a}}^t$. Finally, overlapping parts of the corresponding sub-solutions $\hat{\boldsymbol{f}}^t = \boldsymbol{D}\hat{\boldsymbol{a}}^t$ are averaged in order to obtain the reconstructed signal $\hat{\boldsymbol{f}}$.

It is assumed that the sub-signals have sparse representations in terms of the discrete cosine basis, i.e., that the discrete cosine transforms $\mathrm{DCT}(\hat{\boldsymbol{f}}^t) = \hat{\boldsymbol{a}}^t$ are sparse. As a consequence, one obtains that $\hat{\boldsymbol{f}}^t = \mathrm{IDCT}(\hat{\boldsymbol{a}}^t)$ and hence, $\boldsymbol{D} = \mathrm{IDCT}(\boldsymbol{I}_n)$ is used in (6.9) and (6.15). In order to split $\boldsymbol{q}$, the size $n \leq N$ of the sub-signals as well as a shift length $s \leq n$ are fixed. Therewith, the t-th sub-signal is given by

$$\boldsymbol{q}^t = (q_{(t-1)s+1}, \ldots, q_{(t-1)s+n}). \tag{6.17}$$

Considering that necessarily $(t-1)s+n \leq N$, we obtain sub-signals for $t = 1, \ldots, \lfloor \frac{N-n}{s} \rfloor + 1$.

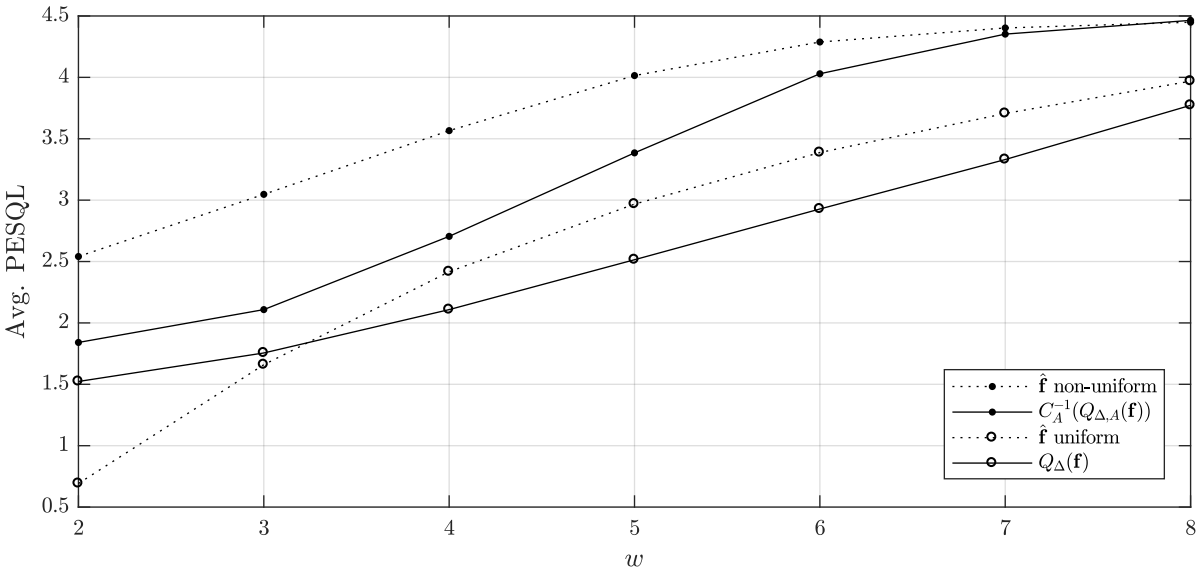

Figure 6.8: Average PESQL values obtained in experiments with 720 speech signals from the IEEE corpus using $n = 1024$, $s = 256$, word lengths $w = 2, \ldots, 8$ and 25 iterations in Chambolle-Pock's algorithm, compared to average PESQL values of the standard reconstructions $C_A^{-1}(Q_{\Delta,A}(\boldsymbol{f}))$ and $Q_\Delta(\boldsymbol{f})$, respectively.

To evaluate the speech quality of the reconstructed signal, the authors of [4] use the *perceptual evaluation of speech quality* (PESQ) measure (see [27]). The PESQ measure is a so called *full reference* algorithm (i.e., it has access to the reconstruction *and* to the clean signal) consisting of multiple computation stages. Unlike, e.g., the *mean squared error* and the *signal-to-noise ratio*, the PESQ measure aims at expressing the speech quality as it is perceived by a human listener in terms of a number between -0.5 and 4.5 (a higher value stands for a better speech quality). The authors of [4] use the MATLAB implementation provided by [27] which they abbreviate as PESQL.

Figure 6.8 displays average results over 720 male speech signals from the IEEE speech database. The experiments outlined above were performed using $n = 1024$, $s = 256$, word lengths $w = 2, \ldots, 8$ and a fixed number of 25 iterations in Chambolle-Pock's algorithm. Limiting the number of iterations like this leads to fast computational scheme which appears to be convenient for real-time application. In general, 25 iterations are naturally not sufficient to solve the problems (6.9) and (6.15) exactly. However, the experimental results in [4] indicate that 25 iterations are often enough to obtain a reconstruction with a remarkably higher PESQL value compared to the standard reconstructions $Q_\Delta(\boldsymbol{f})$ and $C_A^{-1}(Q_{\Delta,A}(\boldsymbol{f}))$ for uniform and non-uniform quantization, respectively. Moreover, the authors of [4] found that using 50 or more iterations only leads to minor improvement of the average PESQL values, or even to decreasing PESQL values. As Chambolle-Pock's algorithm is proven to converge to an optimal solution of (6.9) and (6.15), respectively, this observation indicates that the exact solutions of the respective problems are apparently not optimal in terms of the perceived speech quality.

The observation that optimal solutions of (6.9) and (6.15) are not necessarily good solutions in terms of speech quality limits the applicability of ℓ_1-HOUDINI to speech dequantization. On the one hand, performing only a fixed number of iterations of ℓ_1-HOUDINI does never yield a feasible solution of (6.9) and (6.15) (except if the solution is already optimal), whereas a fixed number of iterations of Chambolle-Pock's algorithm with the respective starting points (located inside the feasible region) leads to solutions that are at least approximately feasible. On the other hand, we observed that calculating the full reconstructed signal $\hat{\boldsymbol{f}}$ by applying ℓ_1-HOUDINI to the above-mentioned sub-problems requires computational running times that are several magnitudes higher than the actual length of the signal (in seconds). Hence, using ℓ_1-HOUDINI in the context of speech dequantization does not qualify for any real-time application. The same statement holds true if we use GUROBI to solve the LP reformulations of the subproblems or perform Chambolle-Pock iterations until the algorithm has converged.

In the case of ℓ_1-HOUDINI, one particular difficulty concerning the initialization $\boldsymbol{x}^0 = \boldsymbol{0}$ and the according active set $W = \{i : |b_i| = \|\boldsymbol{b}\|_\infty\}$ (see Section 3.4) showed up during our numerical experiments with speech dequantization. Namely, as in this application the vector $\boldsymbol{b}$ substantially depends on the quantized signal, which can only take a small number of different values, the number of elements in W is likely to be large. In turn, the potential number of different support sets $\Omega \subseteq W$ of the first dual iterate $\boldsymbol{y}^1$ is then as well large. As a consequence our active-set implementation for the dual update struggled to identify the optimal dual support, although it finally terminated at an optimal point.

The circumstance that a truncated algorithmic scheme for (6.9) and (6.15) apparently fares better than an exact solver for the respective problems raises the question whether we can adapt our model such that the associated optimal solutions are in some sense similar to the outcome of the truncated scheme. To get an intuition about the differences between both approaches, we illustrate some examples in Figures 6.9–6.11.

Figures 6.9 and 6.10 show a speech signal which has, except one single peak, relatively low magnitude. In all plots, the grid lines correspond with the uniform and non-uniform quantization intervals, respectively. As expected, the uniformly quantized signal $Q_\Delta(\boldsymbol{f})$ has its values exactly in the middle of the respective quantization intervals, while the components of $C_A^{-1}(Q_{\Delta,A}(\boldsymbol{f}))$ are located in the interior (but not in the middle) of the non-uniform quantization intervals. It can be observed that, due to finer quantization of the lower magnitudes, the exact solution of (6.15) approximates the original signal far better than the exact solution of (6.9). However, the exact solutions are in both cases obviously clipped to the boundaries of the surrounding quantization intervals. This behavior is a result of the optimality conditions for (P_δ) which require that $|\boldsymbol{a}_i^\top \boldsymbol{x}^\star - b_i| = \delta$ whenever the i-th component of the dual solution is non-zero (see Section 3.1). It is obvious that this extent of clipping does not manifest in the case of natural human speech. Moreover, the fact that the approximate reconstructions obtained after 25 iterations of Chambolle-Pock's algorithm are less clipped is likely to explain the associated higher PESQL values. In Figures 6.11 and 6.12, the clipping effect is clearly visible, particularly in view of the non-uniformly reconstructed signal. Here, the uniformly

reconstructed signal is actually relatively closer to the original signal because most components have rather high magnitude and non-uniform quantization does therefore not pay off.

To conclude this section, we show a statistically motivated approach by which we can represent our previous model, but which allows us to incorporate additional information as well, e.g., that the reconstructed signals should largely not be clipped to the boundaries of the quantization intervals.

6.1.6 MAP Estimation

Previously, our approach was to reconstruct a speech signal $\boldsymbol{f}$ on the basis of quantized measurements $\boldsymbol{q} = Q(\boldsymbol{f})$ assuming that

1. the reconstructed signal satisfies $\hat{\boldsymbol{f}} \in Q^{-1}(\boldsymbol{q})$ and
2. there exist a sparse coefficient vector and a known dictionary such that $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{a}$.

As a consequence, we established the optimization problem (6.6) in order to find an approximation of the coeffiecient vector. In the following, we show that each of the above assumptions can be modeled in terms of a respective *probability mass function* (pmf) or *probability density function* (pdf), respectively, such that the solution $\hat{\boldsymbol{a}}$ in (6.6) coincides with the related *maximum a posteriori* (MAP) estimator.

We use the first one of the above assumptions in combination with the representation $\boldsymbol{f} = \boldsymbol{D}\boldsymbol{a}$ to model the *likelihood* of $\boldsymbol{q}$ on $\boldsymbol{a}$ in terms of the pmf

$$p(\boldsymbol{q}|\boldsymbol{a}) \propto \mathbf{1}_{Q^{-1}(\boldsymbol{q})}(\boldsymbol{D}\boldsymbol{a}), \tag{6.18}$$

i.e., the probability of observing $\boldsymbol{q}$ given that $\boldsymbol{a}$ is the true coefficient vector is non-zero if and only if $\boldsymbol{D}\boldsymbol{a} \in Q^{-1}(\boldsymbol{q})$. Note that we can a use a pmf because $\boldsymbol{q}$ can only take finitely many values.

The assumption that $\boldsymbol{a}$ is sparse in combination with the fact that minimal ℓ_1-norm solutions of linear equation systems tend to be the sparsest solutions as well (see [17]) leads us to modeling the *prior probability* of $\boldsymbol{a}$ in the form of

$$p(\boldsymbol{a}) \propto \boldsymbol{e}^{-\|\boldsymbol{a}\|_1}. \tag{6.19}$$

Hence, relatively high probability is assigned to subsets of vectors with small ℓ_1-norm, whereas the prior pdf tends to zero with increasing ℓ_1-norm of $\boldsymbol{a}$.

By Bayes' Theorem (see [32]), the *posterior probability* of $\boldsymbol{a}$ is given in terms of the pdf

$$p(\boldsymbol{a}|\boldsymbol{q}) = \frac{p(\boldsymbol{q}|\boldsymbol{a})p(\boldsymbol{a})}{\int_{\mathbb{R}^n} p(\boldsymbol{q}|\tilde{\boldsymbol{a}})p(\tilde{\boldsymbol{a}})\mathrm{d}\tilde{\boldsymbol{a}}}. \tag{6.20}$$

By means of the likelihood and the prior probability, the posterior probability incorporates our entire knowledge about the sought coefficient vector $\boldsymbol{a}$. The *maximum a*

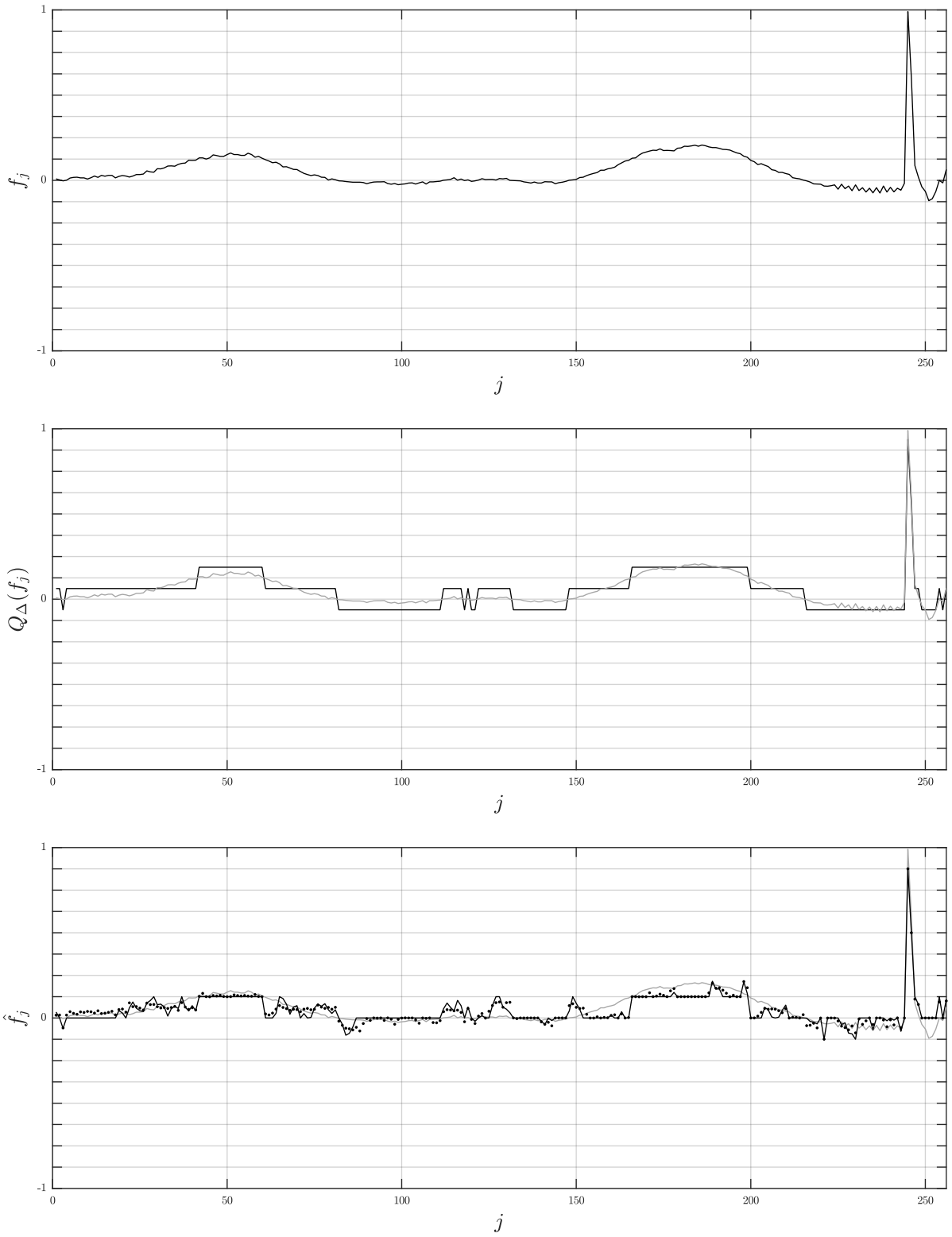

Figure 6.9: Top: Low-magnitude speech signal snippet with $n = 256$ sampling points. Middle: Uniformly quantized speech signal at data rate $w = 4$ (bold line) and original signal (shaded line). Bottom: Optimal solution of (6.9) (bold line), 25th iterate of Chambolle-Pock's algorithm (dotted line) and original signal (shaded line).

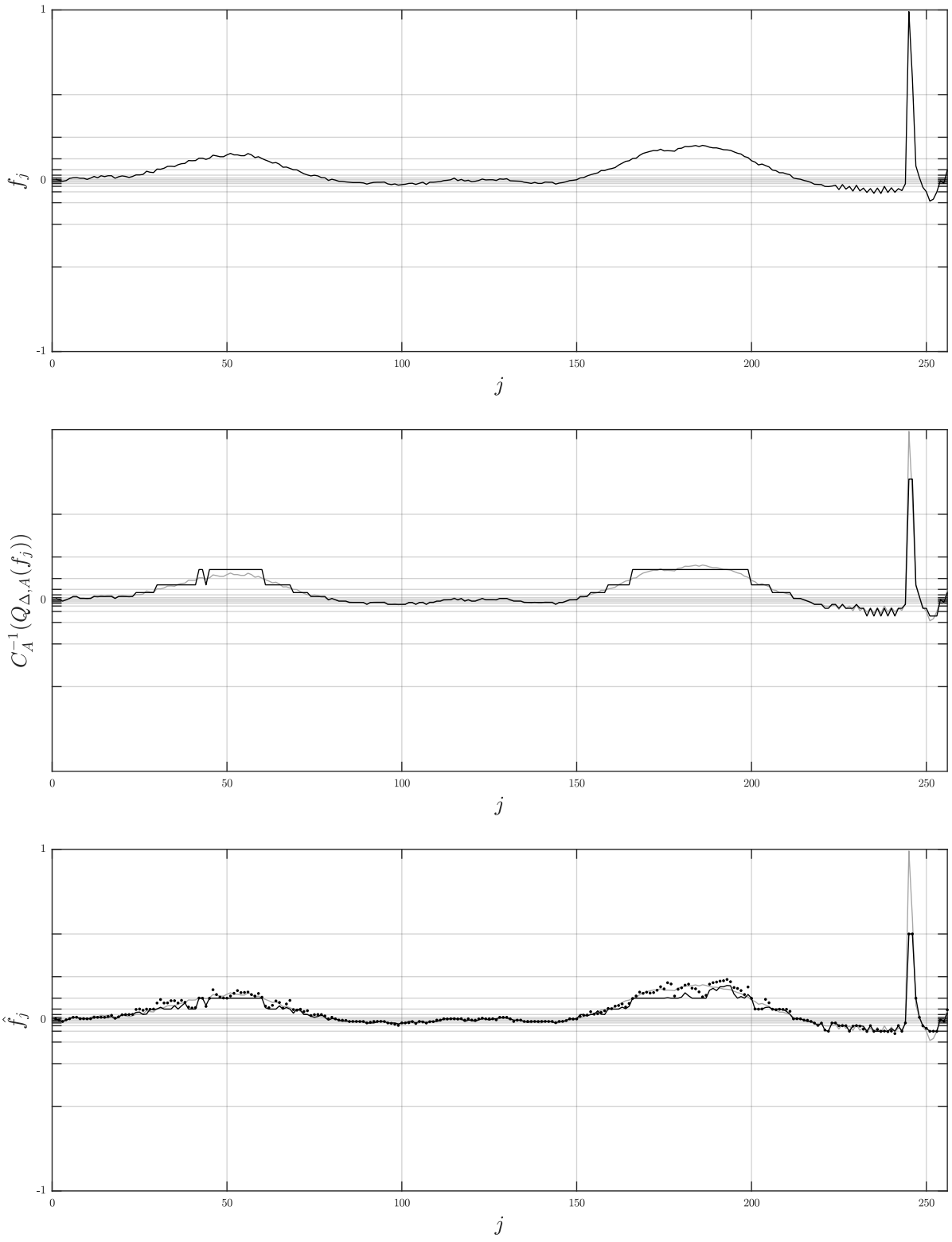

Figure 6.10: Top: Low-magnitude speech signal snippet with $n = 256$ sampling points. Middle: Non-uniformly quantized speech signal at data rate $w = 4$ and with $A = 87.6$ (bold line) and original signal (shaded line). Bottom: Optimal solution of (6.15) (bold line), 25th iterate of Chambolle-Pock's algorithm (dotted line) and original signal (shaded line).

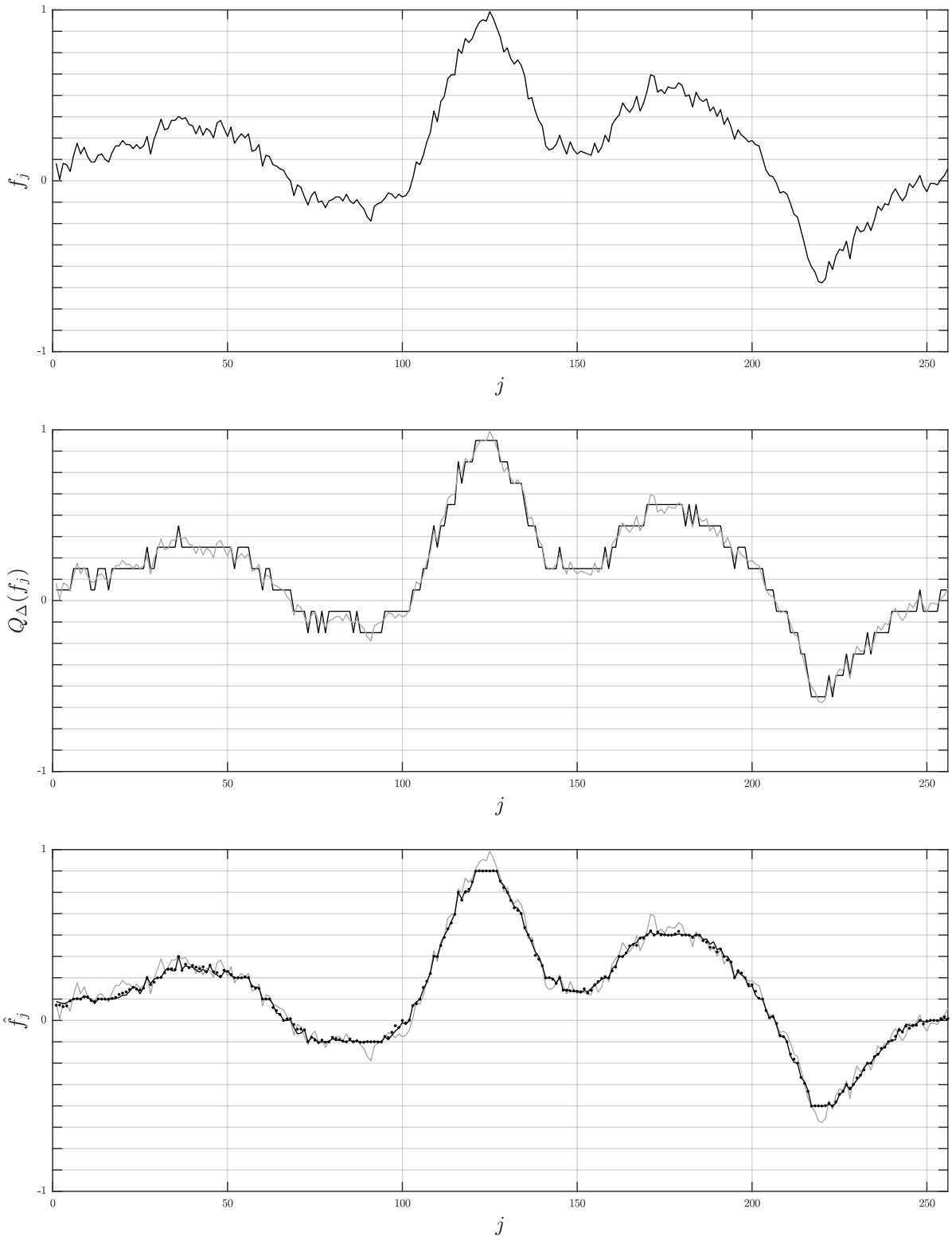

Figure 6.11: Top: High-magnitude speech signal snippet with $n = 256$ sampling points. Middle: Uniformly quantized speech signal at data rate $w = 4$ (bold line) and original signal (shaded line). Bottom: Optimal solution of (6.9) (bold line), 25th iterate of Chambolle-Pock's algorithm (dotted line) and original signal (shaded line).

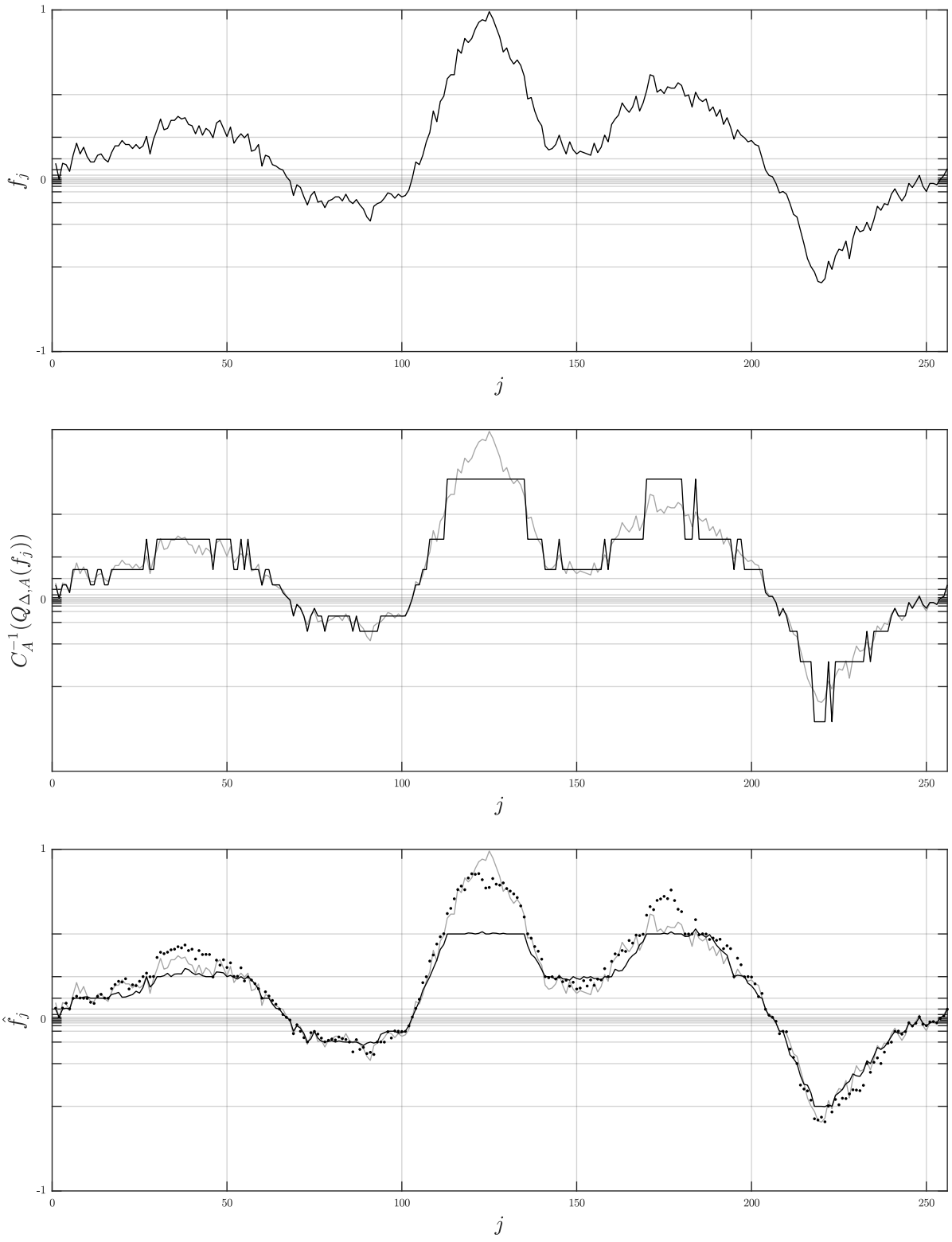

Figure 6.12: Top: High-magnitude speech signal snippet with $n = 256$ sampling points. Middle: Non-uniformly quantized speech signal at data rate $w = 4$ and with $A = 87.6$ (bold line) and original signal (shaded line). Bottom: Optimal solution of (6.15) (bold line), 25th iterate of Chambolle-Pock's algorithm (dotted line) and original signal (shaded line).

posteriori (MAP) estimate (also called the *posterior mode*, see [32]) is a point estimate which maximizes the posterior pdf, i.e.,

$$\boldsymbol{a}_{\text{MAP}} \in \underset{\boldsymbol{a}\in\mathbb{R}^n}{\arg\max}\ p(\boldsymbol{a}|\boldsymbol{q}). \tag{6.21}$$

Since the denominator in (6.20) does not depend on $\boldsymbol{a}$, the MAP estimate can be calculated according to

$$\boldsymbol{a}_{\text{MAP}} \in \underset{\boldsymbol{a}\in\mathbb{R}^n}{\arg\max}\ p(\boldsymbol{q}|\boldsymbol{a})p(\boldsymbol{a}) = \underset{\boldsymbol{a}\in\mathbb{R}^n}{\arg\max}\ \mathbf{1}_{Q^{-1}(\boldsymbol{q})}(\boldsymbol{D}\boldsymbol{a})e^{-\|\boldsymbol{a}\|_1}. \tag{6.22}$$

As the objective function in (6.22) is non-negative, the maximum will clearly be attained at some point satisfying $\boldsymbol{D}\boldsymbol{a} \in Q^{-1}(\boldsymbol{q})$. Moreover, it is equivalent to minimize $\|\boldsymbol{a}\|_1$ instead of maximizing $\exp(-\|\boldsymbol{a}\|_1)$. Therefore, we finally see that

$$\boldsymbol{a}_{\text{MAP}} \in \underset{\boldsymbol{a}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{a}\|_1 \quad \text{s.t. } \boldsymbol{D}\boldsymbol{a} \in Q^{-1}(\boldsymbol{q}). \tag{6.23}$$

All in all, we have shown that there exists a statistical model such that our previously derived approximation $\hat{\boldsymbol{a}}$ is related to the corresponding MAP estimate $\boldsymbol{a}_{\text{MAP}}$. Hence, the described decoding procedures have interpretations as statistical-model-based algorithms in the sense of [27]. Although we have yet only reproduced our earlier results using a statistically motivated approach, this interpretation opens the door to refined models which incorporate, e.g., empirical information about the distribution of the coefficient vectors. More precisely, we could model the prior probability of $\boldsymbol{a}$ using a pdf

$$p(\boldsymbol{a}) \propto f(\boldsymbol{a})e^{-\|\boldsymbol{a}\|_1}, \tag{6.24}$$

where the function f introduces the additional information about $\boldsymbol{a}$. With regard to the MAP estimator, we then obtain

$$\boldsymbol{a}_{\text{MAP}} \in \underset{\boldsymbol{a}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{a}\|_1 - \ln f(\boldsymbol{a}) \quad \text{s.t. } \boldsymbol{D}\boldsymbol{a} \in Q^{-1}(\boldsymbol{q}). \tag{6.25}$$

However, we do not delve into statistical models for speech dequantization here and leave this aspect open as a possible subject of future work.

6.2 The Dantzig Selector

Suppose that $\boldsymbol{y} \in \mathbb{R}^n$ is a vector of *observations*, that $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is a *predictor matrix* and that $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$ is a vector of independent and identically distributed *stochastic measurement errors*. The problem of estimating a parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ from the linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{z}$ arises in numerous applications. The authors of [11] consider the specific situation when p is much larger than n which is, e.g., related to applications in radiology and biomedical imaging where one has often far fewer measurements than unknown parameters.

6.2.1 Optimization Problem and Motivation

As an estimate of the true paramter vector, the authors of [11] propose to choose

$$\hat{\boldsymbol{\beta}} \in \underset{\tilde{\boldsymbol{\beta}}\in\mathbb{R}^p}{\arg\min} \|\tilde{\boldsymbol{\beta}}\|_1 \quad \text{s.t. } \|\boldsymbol{X}^\top(\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y})\|_\infty \leq \lambda_p\sigma. \tag{6.26}$$

While (6.26) is clearly a special case of (P_δ), one particular aspect of (6.26) is that the product $\boldsymbol{X}^\top(\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y})$ is constrained rather than only $\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y}$. According to [11] there are, besides good theoretical approximation guarantees, at least two intuitive reasons why this can be of advantage. First, the estimation scheme (6.26) is invariant to orthogonal transformations, i.e., for any orthogonal matrix $\boldsymbol{U}\in\mathbb{R}^{n\times n}$ it holds that

$$(\boldsymbol{U}\boldsymbol{X})^\top(\boldsymbol{U}\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{U}\boldsymbol{y}) = \boldsymbol{X}^\top(\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y}). \tag{6.27}$$

Second, consider an example with $\sigma > 1/\sqrt{n}$ and $\lambda_n = \sqrt{2\ln n}$.[1] Then, it holds for $n>1$ that $1/\sqrt{n} \leq \lambda_n\sigma < 1$. Further, suppose that $\boldsymbol{r} := \boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y} = \boldsymbol{X}^i$ holds for the i-th column with $|\boldsymbol{X}^i| = \mathbb{1}/\sqrt{n}$ and hence, $\|\boldsymbol{r}\|_\infty \leq \lambda_n\sigma$. Although it makes no sense that $\tilde{\boldsymbol{\beta}}$ is a feasible solution, since with $\boldsymbol{X}(\tilde{\boldsymbol{\beta}}-\boldsymbol{e}_i)-\boldsymbol{y}=\boldsymbol{0}$ the i-th variable is rightly included into the model providing an exact data fit, $\tilde{\boldsymbol{\beta}}$ is feasible for the problem where only the ℓ_∞-norm of $\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y}$ is constrained. However, $\tilde{\boldsymbol{\beta}}$ is not feasible for (6.26) because it holds that $\|\boldsymbol{X}^\top\boldsymbol{r}\|_\infty \geq 1$. In short: the constraint in (6.26) prevents the residual $\boldsymbol{r}$ from being too correlated with the columns of $\boldsymbol{X}$.

6.2.2 Algorithmic Approaches

The authors of [11] propose to solve the LP reformulation

$$\min_{\boldsymbol{u},\tilde{\boldsymbol{\beta}}\in\mathbb{R}^p} \mathbb{1}^\top\boldsymbol{u} \quad \text{s.t. } -\boldsymbol{u} \leq \tilde{\boldsymbol{\beta}} \leq \boldsymbol{u}, \quad -\lambda_p\sigma\mathbb{1} \leq \boldsymbol{X}^\top(\boldsymbol{X}\tilde{\boldsymbol{\beta}}-\boldsymbol{y}) \leq \lambda_p\sigma\mathbb{1} \tag{6.28}$$

of (6.26) using a specific primal-dual interior-point algorithm (an implementation is part of the collection ℓ_1-Magic, see [9]). In addition, a dedicated homotopy method for the Dantzig selector named *Primal Dual pursuit* (PDP) has been introduced in [1]. As both ℓ_1-Houdini and PDP are homotopy algorithms, we continue with a short comparison of both algorithms, before we proceed with some numerical results.

The basic idea behind both ℓ_1-Houdini and PDP is that primal and dual update steps are derived based on primal-dual optimality conditions. Apart from this, there are some significant differences between both algorithms. First, PDP is specialized to the Dantzig selector problem with constraints $\|\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\|_\infty \leq \delta$, whereas ℓ_1-Houdini can handle more general constraints in the form of $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_\infty \leq \delta$. Second, PDP performs explicit updates of the primal and dual variables requiring the existence of the inverses $(\boldsymbol{A}_\Omega^\top\boldsymbol{A}_S)^{-1}$ and $(\boldsymbol{A}_S^\top\boldsymbol{A}_\Omega)^{-1}$ in each iteration. In contrast, ℓ_1-Houdini updates the primal and dual iterates via linear programming without needing the above-mentioned matrices

[1] This value of λ_n corresponds to the choice in [11, Theorem 1.1] with $n=p$.

to be invertible. Third, PDP identifies $S = \Sigma$ and $\Omega = W$ and each set changes in at most one index per iteration. As a consequence of the second aspect above, it further holds that $|S| = |W| = |\Omega| = |\Sigma|$ after each iteration of PDP. However, ℓ_1-HOUDINI explicitly allows for proper subsets $S \subset \Sigma$ and $\Omega \subset W$, and multiple changes in the respective sets are possible in each iteration.

6.2.3 Numerical Experiments

Table 6.1: Runtime and accuracy comparison for the Dantzig selector.

inst.	runtime in seconds			$\|\hat{\boldsymbol{\beta}}\|_1$			constraint violation		
	HOU	PDP	GUR	HOU	PDP	GUR	HOU	PDP	GUR
1	0.19	0.14	2.22	97.09	97.09	97.09	$3{\cdot}10^{-15}$	$4{\cdot}10^{-15}$	$3{\cdot}10^{-15}$
2	1.02	0.64	2.36	154.93	154.93	154.93	$3{\cdot}10^{-15}$	$7{\cdot}10^{-15}$	$4{\cdot}10^{-15}$
3	0.34	0.27	8.93	96.41	96.41	96.41	$3{\cdot}10^{-15}$	$3{\cdot}10^{-15}$	$4{\cdot}10^{-15}$
4	2.74	1.48	9.19	188.03	188.03	188.03	$4{\cdot}10^{-15}$	$1{\cdot}10^{-14}$	$6{\cdot}10^{-15}$
5	0.21	0.26	2.26	98.68	98.68	98.68	$3{\cdot}10^{-15}$	$5{\cdot}10^{-15}$	$2{\cdot}10^{-15}$
6	0.47	0.52	2.35	152.03	152.03	152.03	$5{\cdot}10^{-15}$	$1{\cdot}10^{-14}$	$5{\cdot}10^{-15}$
7	0.44	0.41	9.11	95.73	95.73	95.73	$5{\cdot}10^{-15}$	$6{\cdot}10^{-15}$	$5{\cdot}10^{-15}$
8	0.84	0.86	9.22	186.19	186.19	186.19	$5{\cdot}10^{-15}$	$1{\cdot}10^{-14}$	$5{\cdot}10^{-15}$
9	0.03	0.02	< 0.01	44.64	44.64	9.36	$3{\cdot}10^{-10}$	$3 \cdot 10^{-4}$	$2 \cdot 10^{-2}$
10	0.03	0.02	< 0.01	304.27	304.27	6.03	$1 \cdot 10^{-8}$	$4 \cdot 10^{-3}$	$2 \cdot 10^{-1}$
11	0.02	0.01	< 0.01	316.35	316.35	316.35	$7 \cdot 10^{-8}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-7}$
12	0.04	0.02	< 0.01	64.18	64.18	64.18	$3 \cdot 10^{-9}$	$6 \cdot 10^{-7}$	$7{\cdot}10^{-10}$
13	0.02	-	0.03	0.79	-	$2 \cdot 10^5$	$7 \cdot 10^{-7}$	-	$4 \cdot 10^{-9}$
14	0.21	3.47	0.52	0.67	1.88	634.89	$7 \cdot 10^{-7}$	$1 \cdot 10^{-7}$	$1{\cdot}10^{-11}$
15	176.76	5.52	1.11	998.72	157.41	998.72	$8 \cdot 10^{-7}$	$4 \cdot 10^4$	$4 \cdot 10^{-7}$

We compare our method to PDP and to the commercial LP solver GUROBI, where we apply the latter to the LP reformulation of (6.26) according to (5.2). Our test set includes random instances and several instances from [26]. Table 6.2 provides an overview of the instances and Table 6.1 summarizes the numerical results. To generate the random instances, we adopt the following procedure from [11]: First, we generate $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ with independent Gaussian entries and afterwards normalize all columns such that $\|\boldsymbol{X}^i\|_2 = 1$ holds. Then, we choose $\boldsymbol{\beta} \in \mathbb{R}^p$ at random with a certain sparsity $|S|$, fix $\sigma \coloneqq \sqrt{|S|/n}/3$ and set $\boldsymbol{y} \coloneqq \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{z}$ with $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. Finally, we determine λ_p by taking the maximum of $\|\boldsymbol{X}^\top \tilde{\boldsymbol{z}}\|_\infty$ over 100 realizations of $\tilde{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.

All experiments were conducted on Ubuntu with an Intel® Core™ i7-4550U CPU @ 1.50GHz × 4 processor), using MATLAB R2017a, the dual simplex solver of GUROBI 7.5.1 (via its MATLAB interface), the PDP implementation from the ℓ_1-HOMOTOPY package [2] and ℓ_1-HOUDINI in combination with the active-set method described in Chapter 4.

Table 6.2: Test instances for the Dantzig selector.

inst.	description	n	p	$\lambda_p \sigma$	$\lvert S \rvert$
1	random [11]	1024	1024	0.39	66
2	random [11]	1024	1024	0.51	152
3	random [11]	1024	2048	0.27	69
4	random [11]	1024	2048	0.39	166
5	random [11]	2048	1024	0.35	65
6	random [11]	2048	1024	0.55	128
7	random [11]	2048	2048	0.29	64
8	random [11]	2048	2048	0.39	130
9	Wine (red) [14, 26]	1599	12	0.00	12
10	Wine (white) [14, 26]	4898	12	0.00	12
11	Airfoil Self-Noise [30, 26]	1503	6	0.00	6
12	Housing [23, 26]	506	14	0.00	14
13	Online News Popularity [19, 26]	39644	59	0.00	6
14	Blog Feedback [6, 26]	52396	280	0.00	11
15	Relative location of CT sclices on axial axis [21, 22, 26]	53500	385	0.00	385

The first part of the comparison in Table 6.1 shows that the runtimes of ℓ_1-Houdini and PDP often lie in the same magnitude while the respective runtimes of Gurobi are significantly larger. We can further observe that ℓ_1-Houdini is fastest in case $n > p$ which is of interest in many machine learning applications, where the number of training examples is much larger that the number of features. Applied to the empirical data from [26], Gurobi is the fastest algorithm in the majority of cases, while PDP fails to find an optimal solution in more than one case (see instances 13 and 15). Table 6.1 finally shows that ℓ_1-Houdini is the only algorithm that works with high accuracy on the whole test set. Note that our experiments also included ℓ_1-Magic applied to the LP formulation (6.28). We omitted the corresponding results in Table 6.1 because ℓ_1-Magic performed worse than all remaining methods on the vast majority of test instances.

6.3 Sparse Precision Matrix Estimation

Estimating the *covariance matrix* and its inverse (the *precision matrix*) based on a sample from some distribution is an important problem in various statistical applications. In the following, we introduce an approach for *sparse precision matrix estimation* which was proposed in [8], and which is another example for the applicability of ℓ_1-Houdini.

Let $\boldsymbol{X} \in \mathbb{R}^p$ be a random vector with covariance matrix $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{p \times p}$ and precision matrix $\boldsymbol{\Omega}_0 \coloneqq \boldsymbol{\Sigma}_0^{-1}$, and suppose that both matrices are unknown. Further let $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ be an independent and identically distributed sample from the distribution of $\boldsymbol{X}$. A

common estimator for the covariance matrix is

$$\boldsymbol{\Sigma}_n \coloneqq n^{-1} \sum_{k=1}^{n} (\boldsymbol{X}_k - \bar{\boldsymbol{X}})(\boldsymbol{X}_k - \bar{\boldsymbol{X}})^\top, \tag{6.29}$$

where

$$\bar{\boldsymbol{X}} \coloneqq n^{-1} \sum_{k=1}^{n} \boldsymbol{X}_k \tag{6.30}$$

is the mean of the sample. In case an estimate of the precision matrix is of particular interest, one could be tempted to determine the inverse of $\boldsymbol{\Sigma}_n$. However, this matrix is singular in case $n < p$ and hence, $\boldsymbol{\Sigma}_n^{-1}$ is not a well-defined estimator for $\boldsymbol{\Omega}_0$ in general. To overcome this issue, the authors of [8] propose to determine

$$\hat{\boldsymbol{\Omega}}_1 \;\coloneqq\; \operatorname*{arg\,min}_{\boldsymbol{\Omega} \in \mathbb{R}^{p \times p}} \; \|\boldsymbol{\Omega}\|_1 \quad \text{s.t. } \|\boldsymbol{\Sigma}_n \boldsymbol{\Omega} - \boldsymbol{I}_p\|_\infty \leq \lambda \tag{6.31}$$

in a first step, where $\lambda > 0$ is some tuning parameter. Afterwards, a symmetrized version $\hat{\boldsymbol{\Omega}}$ of $\hat{\boldsymbol{\Omega}}_1$ is taken as an estimate of the precision matrix. The authors of [8] call this the CLIME estimator.

The problem (6.31) naturally decomposes into p vector valued problems with p variables. More precisely, if $\boldsymbol{e}_i$ is the i-th standard unit vector and

$$\hat{\boldsymbol{\beta}}_i \;\coloneqq\; \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^p} \; \|\boldsymbol{\beta}\|_1 \quad \text{s.t. } \|\boldsymbol{\Sigma}_n \boldsymbol{\beta} - \boldsymbol{e}_i\|_\infty \leq \lambda, \tag{6.32}$$

then $\hat{\boldsymbol{\Omega}}_1 = [\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_p]$ is a solution of (6.31). The problem (6.32) has again the form of (P_δ) and can thus be solved using ℓ_1-Houdini.

6.4 Sparse Linear Discriminant Analysis

The classification of high-dimensional data is a recent problem, for instance in machine learning. Given two p-variate normal distributions $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ (class 1) and $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ (class 2) with the same covariance matrix as well as a random vector $\boldsymbol{Z}$ drawn from one of these distributions, the goal of classification is to decide from which of the two distributions $\boldsymbol{Z}$ is drawn. With $\boldsymbol{\mu} \coloneqq (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2$, $\boldsymbol{\delta} \coloneqq \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ and $\boldsymbol{\Omega} \coloneqq \boldsymbol{\Sigma}^{-1}$ (the precision matrix), *Fisher's linear discriminant rule*

$$\psi_{\mathbf{F}}(\boldsymbol{Z}) = \mathbf{1}_{\mathbb{R}_+^0}([\boldsymbol{Z} - \boldsymbol{\mu}]^\top \boldsymbol{\Omega} \boldsymbol{\delta}) \tag{6.33}$$

classifies $\boldsymbol{Z}$ into class 1 if and only if $\psi_{\mathbf{F}}(\boldsymbol{Z}) = 1$.

As the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are typically unknown, it is often not possible to apply (6.33) directly. A standard approach is to estimate the parameters on the basis of samples $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}\}$ and $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}\}$, i.e., to use the means

$$\bar{\boldsymbol{X}} \coloneqq n_1^{-1} \sum_{k=1}^{n_1} \boldsymbol{X}_k \quad \text{and} \quad \bar{\boldsymbol{Y}} \coloneqq n_2^{-1} \sum_{k=1}^{n_2} \boldsymbol{Y}_k \tag{6.34}$$

as estimates for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and the sample covariance matrix

$$\hat{\boldsymbol{\Sigma}}_n \coloneqq (n_1 + n_2)^{-1} \left[\sum_{k=1}^{n_1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top + \sum_{k=1}^{n_2} (\boldsymbol{Y}_k - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_k - \bar{\boldsymbol{Y}})^\top \right] \tag{6.35}$$

as an estimate for $\boldsymbol{\Sigma}$. Therewith, the inverse $\hat{\boldsymbol{\Sigma}}_n^{-1}$ is the classical estimate for the precision matrix. As mentioned in the previous section, the sample covariance matrix is singular in case $n \coloneqq n_1 + n_2 < p$. Then, one possible approach is to estimate $\boldsymbol{\Omega}$ using the CLIME estimator proposed in [8]. However, the authors of [8] propose a different strategy in [7]. Based on the observation that (6.33) only requires the product $\boldsymbol{\Omega\delta}$, they suggest to estimate this product directly and show that this approach is more effective and efficient than estimating $\boldsymbol{\Omega}$ and $\boldsymbol{\delta}$ separately. To that end, they introduce a sparsity assumption on $\boldsymbol{\Omega\delta}$ and use the estimate

$$\hat{\boldsymbol{\beta}} \;\coloneqq\; \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^p} \; \|\boldsymbol{\beta}\|_1 \quad \text{s.t. } \|\hat{\boldsymbol{\Sigma}}_n \boldsymbol{\beta} - (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}})\|_\infty \leq \lambda_n, \tag{6.36}$$

where λ_n is a tuning parameter. Roughly speaking, the constraint in (6.36) incorporates the assumption that $\hat{\boldsymbol{\Sigma}}_n \boldsymbol{\Omega\delta}$ should be close to $\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}$ which is in turn the natural estimate for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}$. Based on (6.33) and (6.36), and with $\hat{\boldsymbol{\mu}} \coloneqq (\bar{\boldsymbol{X}} + \bar{\boldsymbol{Y}})/2$, the authors of [7] propose to classify $\boldsymbol{Z}$ to class 1 if and only if

$$(\boldsymbol{Z} - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{\beta}} \geq 0. \tag{6.37}$$

Again, we see that (6.36) has exactly the form of (P_δ) and can thus apply ℓ_1-Houdini to solve the problem. On the other hand, the authors of [7] propose to solve the LP reformulation

$$\min_{\boldsymbol{u}, \boldsymbol{\beta} \in \mathbb{R}^p} \; \mathbb{1}^\top \boldsymbol{u} \quad \text{s.t. } -\boldsymbol{u} \leq \boldsymbol{\beta} \leq \boldsymbol{u}, \quad -\lambda_n \mathbb{1} \leq \hat{\boldsymbol{\Sigma}}_n \boldsymbol{\beta} - (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}) \leq \lambda_n \mathbb{1} \tag{6.38}$$

by applying a primal-dual interior-point algorithm (the one that is proposed in [11]). This approach yields only one estimate $\hat{\boldsymbol{\beta}}$ for one respective value of λ_n, whereas one call of ℓ_1-Houdini yields the whole solution path $\hat{\boldsymbol{\beta}}(\lambda_n)$ for all values $\lambda_n \geq 0$ such that (6.36) has a non-empty feasible set. As we will see in the next section, this property of ℓ_1-Houdini can be of advantage in terms of *model selection*.

6.5 Model Selection

In the previous section, we have referred to λ_n as a *tuning parameter*. As in (6.36), the matrix $\hat{\boldsymbol{\Sigma}}_n$ as well as the vectors $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ solely depend on the given samples, it is natural to choose λ_n such that the related minimizer $\hat{\boldsymbol{\beta}}$ induces, in some sense, the best attainable classifier of the type (6.37). In other words, each value of λ_n induces a certain *model* and our goal is to *select* the best one among all models.

6.5.1 General Cross-Validation Scheme

One very common approach for model selection is *cross-validation* (see [25]). The authors of [7] propose the following N*-fold* cross-validation scheme: As a first step, the samples $Z \coloneqq \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_{n_1}\} \cup \{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{n_2}\}$ are divided into N subsets

$$Z_l \coloneqq \{\boldsymbol{X}_i : i \in H_{1l}\} \cup \{\boldsymbol{Y}_j : j \in H_{2l}\} \tag{6.39}$$

(the folds), where mutually disjoint index sets H_{1l} and H_{2l} are chosen such that

$$\bigcup_{l=1}^{N} H_{1l} = \{1, \dots, n_1\} \quad \text{and} \quad \bigcup_{l=1}^{N} H_{2l} = \{1, \dots, n_2\}\,. \tag{6.40}$$

For a fixed λ, the following procedure is repeated for each $l \in \{1, \dots, N\}$: First, we determine a solution $\hat{\boldsymbol{\beta}}_{l,\lambda}$ of (6.36) using only the samples $Z \setminus Z_l$. Therewith, we apply (6.37) to classify all samples in Z_l and calculate the success rate for the l-th fold (the fraction of correctly classified samples) as

$$CV_{l,\lambda} \coloneqq \frac{|\left\{i \in H_{1l} : (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top \hat{\boldsymbol{\beta}}_{l,\lambda} \geq 0\right\}| + |\left\{j \in H_{2l} : (\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top \hat{\boldsymbol{\beta}}_{l,\lambda} < 0\right\}|}{|H_{1l}| + |H_{2l}|} \tag{6.41}$$

(note that $\hat{\boldsymbol{\mu}}_l$ is calculated with respect to the data $Z \setminus Z_l$). Finally, we define the overall success rate for the parameter λ as

$$CV_\lambda \coloneqq N^{-1} \sum_{l=1}^{N} CV_{l,\lambda}. \tag{6.42}$$

The success rate $CV_\lambda \in [0, 1]$ can be interpreted as an indicator for how well the model associated with λ generalizes to data that is not contained in Z. Accordingly, we choose

$$\lambda_n \in \operatorname*{arg\,max}_{\lambda \in \Lambda} CV_\lambda, \tag{6.43}$$

where $\Lambda \subseteq \mathbb{R}_+^0$ is usually some finite discrete grid. Then, we solve (6.36) again with λ_n, this time using the complete set of samples Z, in order to obtain our final estimate $\hat{\boldsymbol{\beta}}$. Note that the computation of λ_n requires the solutions of $N|\Lambda|$ different problems of the form (6.36). Our next goal is to show that N calls of ℓ_1-HOUDINI are enough in order to determine λ_n, even when Λ is a continuous interval.

6.5.2 Grid Independent Cross-Validation

In the following, we show how the solution paths generated by ℓ_1-HOUDINI (applied to the N different problems (6.36) with samples $Z \setminus Z_l$) can be used to determine

$$\lambda_n \in \operatorname*{arg\,max}_{\lambda \in [\lambda_{\min}, \infty)} CV(\lambda), \tag{6.44}$$

where $\lambda_{\min}$ is the smallest parameter such that the feasible set of (6.36) is non-empty for all $l \in \{1, \dots, N\}$, and CV is now a function that maps each parameter λ to the associated success rate. In (6.44), the interval $[\lambda_{\min}, \infty)$ is chosen maximally with respect to the given partition of the samples into the subsets Z_l. To obtain CV, we first determine N mappings CV_l corresponding to the respective folds, of which we then take the average. We continue by establishing the procedure for the computation of CV_l.

Let $\boldsymbol{\beta}^0, \dots, \boldsymbol{\beta}^K$ and $\lambda^0, \dots, \lambda^K$ be the iterates and respective values of the homotopy parameter generated by ℓ_1-HOUDINI applied to (6.36), where the estimates $\hat{\boldsymbol{\Sigma}}_n$, $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ are calculated with respect to the samples $Z \setminus Z_l$. Moreover, recall that the solution path is piecewise linear and that the linear segments are $\left\{\boldsymbol{\beta}^{k-1} + t(\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}) : t \in [0, 1)\right\}$ for $k = 1, \dots, K$. Based on (6.41), we define the mapping

$$CV_l : [\lambda_{\min}, \infty) \to [0, 1], \quad CV_l(\lambda) \coloneqq CV_{l,\lambda} \tag{6.45}$$

which is piecewise constant due to the piecewise linearity of the solution path in combination with the classification rule (6.37).

Jump discontinuities of CV_l correspond to points along the solution path where either

$$(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^{k-1} + s_i(\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})) = 0 \quad \text{for some } i \in H_{1l} \tag{6.46}$$

or

$$(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^{k-1} + t_j(\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})) = 0 \quad \text{for some } j \in H_{2l}. \tag{6.47}$$

Thus, if

$$s_i \coloneqq -\frac{(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top \boldsymbol{\beta}^{k-1}}{(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})} \in (0, 1) \quad \text{for some } i \in H_{1l} \tag{6.48}$$

or

$$t_j \coloneqq -\frac{(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top \boldsymbol{\beta}^{k-1}}{(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})} \in (0, 1) \quad \text{for some } j \in H_{2l}, \tag{6.49}$$

then CV_l has jump discontinuities at

$$\lambda_{1i} \coloneqq \lambda^{k-1} + s_i(\lambda^k - \lambda^{k-1}) \tag{6.50}$$

or

$$\lambda_{2j} \coloneqq \lambda^{k-1} + t_j(\lambda^k - \lambda^{k-1}), \tag{6.51}$$

respectively. If the sign of $(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})$ is positive, then CV_l increases at λ_{1i}, whereas it decreases in case the sign is negative. Vice versa, if the sign of $(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top (\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})$ is positive, then CV_l decreases at λ_{2j}, while it increases in case the sign is negative.

If we obtain $s_i = 0$, then CV_l can not increase at $\lambda_{1i} = \lambda^{k-1}$, whereas it decreases under the same condition as above. In case $t_j = 0$, the function increases under the same condition as above and can not decrease at $\lambda_{2j} = \lambda^{k-1}$. Note that the cases $s_i = 1$ and $t_j = 1$ do not need to be handled separately because they correspond to $s_i = 0$ and

Input: $\boldsymbol{\beta}^0, \ldots, \boldsymbol{\beta}^K$, $\lambda^0, \ldots, \lambda^K$, Z_l, H_{1l}, H_{2l}, $\hat{\boldsymbol{\mu}}_l$
Output: CV_l

`// Identify discontinuities and jump directions of` CV_l`:`
for $k = 1, \ldots, K$ **do**
  $\boldsymbol{d}^k \leftarrow \boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}$
  for $\forall i \in H_{1l}$ **do**
    $g_i \leftarrow (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_l)^\top \boldsymbol{d}^k$
    $s_i \leftarrow$ step size according to (6.50)
    $c_{1i}^k \leftarrow \mathbf{1}_{\mathbb{R}_+ \times (0,1)}(g_i, s_i) - \mathbf{1}_{\mathbb{R}_- \times [0,1)}(g_i, s_i)$
    $\lambda_{1i}^k \leftarrow \mathbf{1}_{\mathbb{R}\setminus\{0\}}(c_{1i}^k)(\lambda^{k-1} + s_i(\lambda^k - \lambda^{k-1})) + I_{\mathbb{R}\setminus\{0\}}(c_{1i}^k)$
  for $\forall j \in H_{2l}$ **do**
    $h_j \leftarrow (\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_l)^\top \boldsymbol{d}^k$
    $t_j \leftarrow$ step size according to (6.51)
    $c_{2j}^k \leftarrow \mathbf{1}_{\mathbb{R}_- \times [0,1)}(h_j, t_j) - \mathbf{1}_{\mathbb{R}^+ \times (0,1)}(h_j, t_j)$
    $\lambda_{2j}^k \leftarrow \mathbf{1}_{\mathbb{R}\setminus\{0\}}(c_{2j}^k)(\lambda^{k-1} + t_j(\lambda^k - \lambda^{k-1})) + I_{\mathbb{R}\setminus\{0\}}(c_{2j}^k)$
`// Compute` CV_l`:`
for $\forall \lambda \in [\lambda^K, \infty)$ **do**
  $CV_l(\lambda) \leftarrow \frac{|H_{1l}| + \sum_{k=1}^K \left(\sum_{i \in H_{1l}} \mathbf{1}_{\mathbb{R}_+^0}(\lambda_{1i}^k - \lambda) c_{1i}^k + \sum_{j \in H_{2l}} \mathbf{1}_{\mathbb{R}_+^0}(\lambda_{2,j}^k - \lambda) c_{2,j}^k \right)}{|H_{1l}| + |H_{2l}|}$
return CV_l

Algorithm 4: Grid independent cross-validation.

$t_j = 0$ as soon as we proceed to the next linear segment (except in case of the very last segment, where the cases $s_i = 1$ and $t_j = 1$ need to be included).

Our scheme for the computation of CV_l is illustrated in Algorithm 4. The steps inside the first for loop represent exactly what we have just discussed in view of (6.46)–(6.51), where the values $c_{1i}^k, c_{2j}^k \in \{-1, 0, 1\}$ reflect whether CV_l changes at the respective points and, if it does so, the direction of the jump discontinuity. In the final step, where $CV_l(\lambda)$ is defined, the term $|H_{1l}|$ in the numerator refers to the fact that with $\boldsymbol{\beta}^0 = \mathbf{0}$ in (6.37), all samples $\boldsymbol{X}_i \in H_{1l}$ are classified correctly, while all $\boldsymbol{Y}_j \in H_{2l}$ are assigned to the wrong class. From that point on, we simply add up all jumps of CV_l along the solution path, from λ^0 downwards to λ, in order to obtain $CV_l(\lambda)$.

Referring to the jump discontinuities of CV_l as

$$\lambda_{1i}^{k,l} := \lambda_{1i}^k \quad \text{and} \quad \lambda_{2j}^{k,l} := \lambda_{2j}^k, \tag{6.52}$$

it holds that the function is uniquely determined by $|H_{1l}|$ and the magnitudes of the jumps at the points

$$\Lambda_l := \bigcup_{k=1}^K \left[\bigcup_{i \in H_{1l}} \left\{ \lambda_{1i}^{k,l} \right\} \cup \bigcup_{j \in H_{2l}} \left\{ \lambda_{2j}^{k,l} \right\} \right] \setminus \{\infty\}. \tag{6.53}$$

As we have slightly simplified the calculation, $CV_l(\lambda)$ does not necessarily reflect the true success rate in case $\lambda \in \Lambda_l$. To overcome this issue, we would have had to distinguish between the cases $g_i \gtrless 0$ and $h_j \gtrless 0$, respectively, in order to track whether a jump takes place exactly at or right below the value of λ.

In view of (6.53), it follows that the function

$$CV : [\lambda_{\min}, \infty) \to [0, 1], \quad CV(\lambda) \coloneqq N^{-1} \sum_{l=1}^{N} CV_l(\lambda) \tag{6.54}$$

is as well piecewise constant with jump discontinuities at the points

$$\Lambda \coloneqq \bigcup_{l=1}^{N} \Lambda_l, \tag{6.55}$$

and, as follows from our above discussion, $CV(\lambda)$ almost everywhere reflects the mean success rate over the N folds, except at the points $\lambda \in \Lambda$. However, due to the fact that CV is piecewise constant (on intervals that contain more than a single point each), it holds that

$$\operatorname*{arg\,max}_{\lambda \in [\lambda_{\min}, \infty)} CV(\lambda) = \operatorname*{arg\,max}_{\lambda \in [\lambda_{\min}, \infty) \setminus \Lambda} CV(\lambda). \tag{6.56}$$

If we write $\Lambda = \{\lambda_1, \ldots, \lambda_{|\Lambda|}\}$ with $\lambda_1 < \cdots < \lambda_{|\Lambda|}$ and

$$\lambda_i \in \operatorname*{arg\,max}_{\lambda \in \Lambda} CV(\Lambda), \tag{6.57}$$

then it holds that each model associated with

$$\lambda_n \in (\lambda_{i-1}, \lambda_i) \subseteq \operatorname*{arg\,max}_{\lambda \in [\lambda_{\min}, \infty) \setminus \Lambda} CV(\lambda) \tag{6.58}$$

maximizes the cross-validation succes rate among all attainable models with respect to the chosen partition of Z according to (6.39) and (6.40).

6.5.3 Numerical Experiments

We performed experiments with 10-fold cross-validation on six different randomly generated datasets. While the dimension $p = 500$ is fixed, the total number of samples n passes through $\{100, 200, 300, 400, 500, 600\}$. Throughout, it holds that $n_1 = n_2$ and the number of samples in each fold is $|Z_l| = n/10$. Hence, it holds that $|H_{1l}| = |H_{2l}| = n/20$ for $l = 1, \ldots, 10$. The data were generated at random with $\boldsymbol{\mu}_1 = \mathbb{1}$, $\boldsymbol{\mu}_2 = -\mathbb{1}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p + \boldsymbol{\Upsilon}^\top \boldsymbol{\Upsilon}$, where the entries of $\boldsymbol{\Upsilon} \in \mathbb{R}^{p \times p}$ are independent and identically distributed according to a standard Gaussian distribution.

Figure 6.13 illustrates the mappings $CV : [\lambda_{\min}, \infty) \to [0, 1]$ for each of the above-mentioned experiments and Table 6.3 provides further information on our results. The

values in the second column show that the number of non-zero elements of $\hat{\boldsymbol{\beta}}$ is remarkably close to the number of samples n, as long as it holds that $n \leq p$ (obviously, the number of non-zeros is bounded above by p). Possibly, this can be explained by the observation that, at least in the first four cases where $n < p$, it holds that $\|\hat{\boldsymbol{\beta}}\|_0$ is equal to the rank of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_n$. However, in the remaining cases with $p \leq n$, we observed that $\operatorname{rank}(\hat{\boldsymbol{\Sigma}}_{500}) = 498$ and $\operatorname{rank}(\hat{\boldsymbol{\Sigma}}_{600}) = 500$ which is in both cases greater than $\|\hat{\boldsymbol{\beta}}\|_0$. All in all, as the computational effort to run ℓ_1-HOUDINI tends to increase with the number of non-zeros in the solution, we conclude that it can be particularly beneficial to perform grid independent croos-validation using ℓ_1-HOUDINI in case n is comparably small.

The third column of Table 6.3 shows that, with increasing n, the value of $\lambda_{\min}$ successively decreases until it is finally zero. Apparently, this is the case because $\hat{\boldsymbol{\Sigma}}_n$ becomes more and more regular when we have more samples. However, it does generally not hold that $\lambda_n = \lambda_{\min}$ is an optimal choice according to (6.56), as our experiments with $n \leq 300$ show. There, the optimal value of λ_n is attained on relatively small intervals beyond $\lambda_{\min}$. It is at least not obvious whether we would have found an optimal paramter by using a conventional cross-validation approach on a finite discrete grid. Although we actually do not know how sensitive the success rate on new data (not contained in the set of samples) is to small changes of λ_n, there is no apparent reason not to take the optimal λ_n, even if the optimal success rate is attained on a very small interval. In our experiments with $n \geq 400$, $\lambda_{\min}$ is the left boundary point of the optimal parameter interval. Since $\lambda_{\min}$ and the length of the respective interval are a priori unknown, it is as well not clear whether a conventional cross-validation approach would have yielded an optimal paramter λ_n.

Finally, the last two columns of Table 6.3 give information about the performance of the classification rule (6.37) on the set of samples as well as on newly generated data. The number CV_{sam} represents the success rate on the set of samples. To estimate the success rate on arbitrary data, we generated 10.000 new data points obeying the same distributions as the samples with $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$. After applying (6.37) to the new data, the value CV_{test} reflects the corresponding success rate. The observation that all experiments yield $CV_{\text{test}} \leq CV_{\text{sam}}$ refers to the fact that λ_n and consequently $\hat{\boldsymbol{\beta}}$ are chosen with respect to the set of samples. A large difference between both values, which occurs in particular when the number of samples is relatively small, indicates a certain degree of *overfitting* to the set of samples.

6.6 The L1-Testset

In addition to the experiments outlined above which are all rather application-driven, we performed tests on a subset of instances from the *L1-Testset* described in [28]. This testset includes 548 different instances for the *Basis Pursuit* problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \|\boldsymbol{x}\|_1 \quad \text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{BP}$$

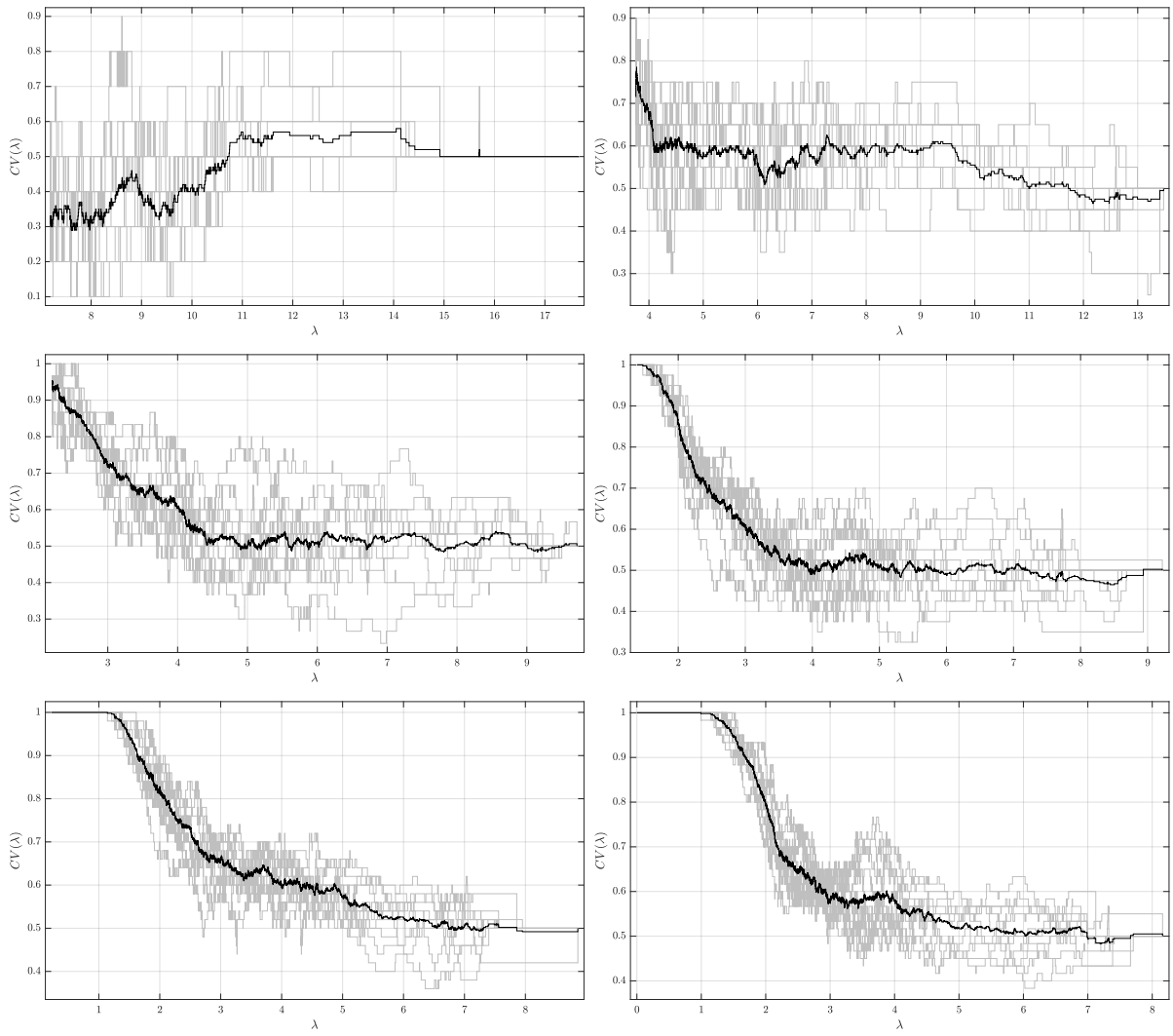

Figure 6.13: Random examples for grid independent 10-fold cross-validation with $p = 500$, $n_1 = n_2$ and $n = n_1 + n_2 \in \{100, 200, 300, 400, 500, 600\}$ (from upper left to lower right). The shaded lines repsresent the functions CV_l for $l = 1, \ldots, 10$ and the bold lines represent the respective means CV.

Table 6.3: Numerical results for grid independent 10-fold cross-validation with varying total numbers of samples n (cf. Figure 6.13).

n	$\|\hat{\boldsymbol{\beta}}\|_0$	$\|\Lambda\|$	$\lambda_{\min}$	λ_n	$CV(\lambda_n)$	CV_{sam}	CV_{test}
100	98	518	7.1929	(14.0507, 14.1437)	0.58	0.68	0.515
200	198	1215	3.7595	(3.77425, 3.77430)	0.7850	0.875	0.5787
300	298	2513	2.2062	$\lambda_{\min}+(2.7, 2.8)\cdot 10^{-6}$	0.9533	0.9733	0.5837
400	398	3154	1.3874	($\lambda_{\min}$, 1.4722)	1	1	0.6146
500	476	3506	0.2227	($\lambda_{\min}$, 1.1420)	1	1	0.8419
600	470	4383	0	($\lambda_{\min}$, 0.9947)	1	1	0.9507

each of which consisting of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, a right-hand side $\boldsymbol{b} \in \mathbb{R}^m$ and an associated optimal solution $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. The instances are constructed such that the *Exact Recovery Condition* (ERC) (see [43]) is satisfied, i.e., $\bar{\boldsymbol{x}}$ is the unique solution of (BP) with matrix $\boldsymbol{A}$ and right-hand side $\boldsymbol{b}$. The following lemma shows that each instance of (BP) with known optimal solution can be used to derive an associated instance of (P_δ) which has the same (not necessarily unique) optimal solution:

Lemma 49. *Let $\bar{\boldsymbol{x}}$ be an optimal solution of* (BP) *with given $\boldsymbol{A}$ and $\boldsymbol{b} = \boldsymbol{A}\bar{\boldsymbol{x}}$. Then, there exists a $\bar{\boldsymbol{y}}$ such that $-\boldsymbol{A}^\top \bar{\boldsymbol{y}} \in \partial\|\bar{\boldsymbol{x}}\|_1$, and for any $\delta > 0$ and $\hat{\boldsymbol{b}} \in \boldsymbol{A}\bar{\boldsymbol{x}} - \delta\partial\|\bar{\boldsymbol{y}}\|_1$, $\bar{\boldsymbol{x}}$ is an optimal solution of* (P_δ) *with the same $\boldsymbol{A}$ and a measurement vector $\boldsymbol{b} = \hat{\boldsymbol{b}}$.*

Proof. It follows from Theorem 12 with $\delta = 0$ that $\bar{\boldsymbol{x}}$ is an optimal solution of (BP) with $\boldsymbol{A}$ and $\boldsymbol{b} = \boldsymbol{A}\bar{\boldsymbol{x}}$ if and only if there exists a vector $\bar{\boldsymbol{y}}$ such that $-\boldsymbol{A}^\top \bar{\boldsymbol{y}} \in \partial\|\bar{\boldsymbol{x}}\|_1$. Choosing $\hat{\boldsymbol{b}} \in \boldsymbol{A}\bar{\boldsymbol{x}} - \delta\partial\|\bar{\boldsymbol{y}}\|_1$, we obtain that $\boldsymbol{A}\bar{\boldsymbol{x}} - \hat{\boldsymbol{b}} \in \delta\partial\|\bar{\boldsymbol{y}}\|_1$. The claim now follows immediately from Theorem 12. □

Following Lemma 49, we first need a dual solution $\bar{\boldsymbol{y}} \in \mathbb{R}^m$ corresponding with the respective instance of (BP) and can then, for arbitrary $\delta > 0$, directly construct a right-hand side $\hat{\boldsymbol{b}}$ such that $\bar{\boldsymbol{x}}$ is an optimal solution of (P_δ) with $\boldsymbol{A}$ and $\hat{\boldsymbol{b}}$. It follows, e.g., from Proposition 31 that the inclusion $-\boldsymbol{A}^\top \bar{\boldsymbol{y}} \in \partial\|\boldsymbol{x}\|_1$ does generally not have a unique solution. In view of the desired instances of (P_δ), this is of interest due to the fact that the active set W associated with $\bar{\boldsymbol{x}}$ is a superset of the support Ω of $\bar{\boldsymbol{y}}$. The authors of [28] describe two ways to solve the inclusion: Either, $\bar{\boldsymbol{y}}$ can be obtained with a closed-form expression or via alternating projections onto $\partial\|\bar{\boldsymbol{x}}\|_1$ and the image space of $\boldsymbol{A}^\top$. However, in our experiments, the accordingly constructed dual solutions $\bar{\boldsymbol{y}}$ were throughout fully dense, i.e., $|\Omega| = m$. To investigate the impact of the optimal primal active set $|W| \geq |\Omega|$ on the performance of the considered solvers, we constructed additional test instances obeying

$$\bar{\boldsymbol{y}} \;\in\; \operatorname*{arg\,min}_{\boldsymbol{y} \in \mathbb{R}^m} \; \|\boldsymbol{y}\|_1 \quad \text{s.t.} \; -\boldsymbol{A}^\top \boldsymbol{y} \in \partial\|\boldsymbol{x}\|_1. \tag{6.59}$$

Note that (6.59) can be recast as a linear program which can be solved efficiently using standard software, even for large-scale problems where an alternating projection approach may no longer work.

In a first set of experiments, we compared the running times of two implementations of ℓ_1-Houdini, one using the active-set approach described in Chapter 4 and one using the commercial LP solver Gurobi to solve the subproblems, to the running times achieved by Gurobi applied to the LP reformulation (5.2) of (P_δ). All three algorithms were applied to a randomly chosen subset of the the L1-Testset with matrix dimensions $512 \times \{1024, 1536, 2048, 4096\}$ and $1024 \times \{2048, 3072, 4096, 8192\}$. For each size, we picked two instances, one in which $\bar{\boldsymbol{x}}$ has non-zero entries of high dynamic range (i.e., the non-zero elements of $\bar{\boldsymbol{x}}$ span several orders of magnitude) and one with low dynamic range. Further, for each of the resulting 16 instances, we constructed one dense dual certificate via alternating projections and one dense dual certificate by solving (6.59).

Hence, our testset finally included 32 different instances, with δ-values drawn randomly from the interval $[0.1, 0.5]$. The experimental results are illustrated in Table 6.4, where the instance numbers refer to the respective indices in the L1-Testset and the running time results were conducted in MATLAB 2016a, using Gurobi 6.5.2 (dual simplex), on Ubuntu with an Intel® Core™ i7-4550U CPU @ 1.50GHz × 4 processor.

In the majority of cases, we observed that ℓ_1-HOUDINI using specialized active-set methods for the subproblems is considerably faster than ℓ_1-HOUDINI using GUROBI (32 out of 32 instances) and even faster than GUROBI used as standalone LP solver (20 out of 32 instances). Another comparison suggests that GUROBI used as standalone solver is usually faster than ℓ_1-HOUDINI using GUROBI for the subproblems (30 out of 32 instances). (Nevertheless, note that ℓ_1-HOUDINI generates the entire solution path with respect to the homotopy parameter, whereas solving the LP formulation of (P_δ) solely yields a solution for the final parameter δ.)

In particular, it seems beneficial to use ℓ_1-HOUDINI when $|S|$ is small (i.e., when the optimal solution $\bar{\boldsymbol{x}}$ is relatively sparse). This is a natural feature of our method since the sparsity of the iterates has direct impact on the size of the subproblems. Analogously, the size of the primal active set W directly affects the size of the subproblems. Our experiments show that solving the very same instance with smaller optimal active set (induced by a modified measurement vector $\hat{\boldsymbol{b}}$) causes an average speedup of 29.6% and 37.6% using ℓ_1-HOUDINI with active-set methods and GUROBI for the subproblems, respectively. In contrast, using GUROBI as standalone LP solver induces an average speedup of 9.4%.

In additional experiments with original instances from the L1-Testset, we observed that ℓ_1-HOUDINI is also competitive in the Basis Pursuit setting ($\delta = 0$). To that end, we compared our method with ℓ_1-HOMOTOPY (see [36], we used the implementation available in [2]), one of the fastest methods according to [28], and again with Gurobi as standalone LP solver. The results are subsumed in Table 6.5. They show that in this special case, ℓ_1-HOUDINI is not as fast as ℓ_1-HOMOTOPY but in most cases still considerably faster than GUROBI.

Table 6.4: Runtime comparison of ℓ_1-HOUDINI against Gurobi. In case of the instances that are marked with a –, the algorithm stopped prematurely because Gurobi failed to solve one of the subproblems.

inst.	$m \times n$	δ	$\lvert\mathcal{S}\rvert$	$\lvert\mathcal{A}\rvert$	ℓ_1-HOUDINI act.-set	ℓ_1-HOUDINI GUROBI	GUROBI standal.
7	512×1024	4.09	34	512	0.98	2.58	0.46
				72	0.53	2.64	0.46
485	512×1024	4.54	51	512	1.80	103.35	1.26
				96	1.22	–	1.09
25	512×1536	0.72	14	512	0.24	3.60	0.83
				31	0.24	3.77	0.81
319	512×1536	4.58	22	512	0.42	15.92	1.63
				43	0.29	10.40	1.53
228	512×2048	3.20	51	512	5.86	–	1.13
				141	3.79	–	0.98
338	512×2048	0.58	20	512	0.79	–	1.93
				45	0.44	16.13	1.42
74	512×4096	1.47	10	512	0.20	18.36	1.26
				38	0.16	1.06	1.22
347	512×2048	2.78	10	512	0.14	8.32	1.25
				32	0.09	0.86	1.21
239	1024×2048	4.79	84	1024	0.77	2.13	0.07
				148	0.82	2.02	0.07
357	1024×2048	4.83	27	1024	1.91	–	3.51
				55	0.80	38.83	2.77
99	1024×3072	0.87	18	1024	0.91	19.65	3.36
				47	0.76	17.49	3.42
527	1024×3072	4.86	99	1024	26.57	–	1.79
				234	16.46	–	1.59
263	1024×4096	4.79	97	1024	36.53	–	2.99
				245	27.36	437.62	2.69
416	1024×4096	2.48	26	1024	2.47	–	6.85
				60	1.33	50.41	3.99
148	1024×8192	4.02	20	1024	1.41	23.21	5.34
				64	1.34	20.67	5.29
421	1024×8192	0.80	9	1024	0.82	–	5.12
				43	0.42	–	5.27

Table 6.5: Runtime comparison of ℓ_1-Houdini (active-set) against ℓ_1-Homotopy (with regularization parameter $\tau = 10^{-9}$) and Gurobi, all applied to the case $\delta = 0$.

inst.	$m \times n$	$\lvert\mathcal{S}\rvert$	ℓ_1-Houdini	ℓ_1-Homotopy	Gurobi
7	512×1024	34	0.83	0.06	0.57
485	512×1024	34	2.09	0.08	1.50
25	512×1536	34	0.28	0.03	0.72
319	512×1536	34	0.47	0.05	1.45
228	512×2048	34	6.37	0.18	0.88
338	512×2048	34	0.90	0.06	2.05
74	512×4096	34	0.23	0.05	1.36
347	512×4096	34	0.16	0.05	1.36
239	1024×2048	34	0.84	0.45	0.08
357	1024×2048	34	2.06	0.11	3.53
99	1024×3072	34	0.94	0.09	3.21
527	1024×3072	34	27.75	0.74	1.83
263	1024×4096	34	35.93	1.08	3.24
416	1024×4096	34	2.53	0.17	9.03
148	1024×8192	34	1.56	0.23	7.23
421	1024×8192	34	0.95	0.15	7.20

7 Conclusion

In this thesis, we have introduced ℓ_1-Houdini, a new homotopy algorithm for ℓ_1-norm minimization with ℓ_∞-norm constraints, as well as a generalized algorithmic scheme which extends the scope of our method to arbitrary linear constraints and, as a consequence, to linear programs with strictly positive objective function coefficients. We have further shown that ℓ_1-Houdini terminates after a finite number of iterations yielding an optimal solution for the problem (P_δ). Subsequently, we have established that ℓ_1-Houdini has to perform at most $(3^{m+n}+1)/2$ iterations in order to find an optimal solution. Afterwards, we have specified a recursive strategy to construct instances of (P_δ), where ℓ_1-Houdini needs to perform exactly $(3^n+1)/2$ iterations in order to find a solution. In diverse examples and numerical experiments, we have demonstrated that our method constitutes an effective, efficient and reliable solver for ℓ_1-norm minimization problems occuring in different fields of application. Besides, we have have described a novel scheme for grid independent cross-validation in the context of sparse linear discriminant analysis, where the availability of the entire solution path of (P_δ), as provided by our method, turns out to be particularly useful.

However, there are a handful of aspects that could not be addressed exhaustively in this thesis. In the first place, this applies to the complexity of our method. We left open the question whether the constructed instances requiring $(3^n+1)/2$ iterations are indeed worst-case examples. If this is the case, then the established upper bound of $(3^{m+n}+1)/2$ iterations is obviously not sharp. In the second place, we have argued that the illustrated model for speech dequantization does apparently not capture the characteristics of human speech really well. Therefore, we have presented an idea how to adapt the model by means of maximum a posteriori estimation, which we did not work out to the end. In the third place, we have not yet investigated applications for our generalized homotopy scheme. Moreover, it is an open question whether it is possible to derive a straightforward variant of ℓ_1-Houdini which is applicable to (P_δ) with an additional convex term in the objective function (instead of or in addition to generalized constraints). Finally, our idea for a grid independent cross-validation scheme does yet only apply to sparse linear discriminant analysis. Thus, the question arises whether the described scheme can be generalized to a broader class of problems where the solution path is available as well.

Bibliography

[1] M. Salman Asif and Justin Romberg. Dantzig selector homotopy with dynamic measurements. In *Proc.SPIE*, volume 72460E, 2009.

[2] M. Salman Asif and Justin Romberg. L1 Homotopy: A MATLAB Toolbox for Homotopy Algorithms in L1 Norm Minimization Problems. http://www.ece.ucr.edu/∼sasif/homotopy, June 2013.

[3] Leonore Blum. A new simple homotopy algorithm for linear programming I. *Journal of Complexity*, 4(4):124–136, June 1988.

[4] Christoph Brauer, Timo Gerkmann, and Dirk Lorenz. Sparse Reconstruction of Quantized Speech Signals. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5940–5944. IEEE, 2016.

[5] Christoph Brauer, Dirk A. Lorenz, and Andreas M. Tillmann. A Primal-Dual Homotopy Algorithm for ℓ_1-Minimization with ℓ_∞-Constraints. *Computational Optimization and Applications*, February 2018.

[6] Krisztian Buza. Feedback Prediction for Blogs. In *Data Analysis, Machine Learning and Knowledge Discovery*, pages 145–152. Springer, 2014.

[7] Tony Cai and Weidong Liu. A Direct Estimation Approach to Sparse Linear Discriminant Analysis. *Journal of the American Statistical Association*, 106(496):1566–1577, December 2011.

[8] Tony Cai, Weidong Liu, and Xi Luo. A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association*, 106(494):594–607, June 2011.

[9] Emmanuel Candès and Justin Romberg. ℓ_1-MAGIC: Recovery of Sparse Signals via Convex Programming. Drawn from https://statweb.stanford.edu/∼candes/l1magic/downloads/l1magic.pdf on November 6, 2017., October 2005.

[10] Emmanuel Candès and Terence Tao. Decoding by Linear Programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, November 2005.

[11] Emmanuel Candès and Terence Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

[12] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[13] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

[14] Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553, 2009.

[15] George B. Dantzig. *Linear Programming and Extensions.* Princeton University Press, 1963.

[16] David L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[17] David L. Donoho. For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ_1-norm Solution Is Also the Sparsest Solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.

[18] Yonina C. Eldar and Gitta Kutinyok, editors. *Compressed Sensing: Theory and Applications.* Cambridge University Press, 2012.

[19] Kelwin Fernandes, Pedro Vinagre, and Paulo Cortez. A Proactive Intelligent Desision Support System for Predicting the Popularity of Online News. In *Proceedings of the 17th EPIA 2015 - Portuguese Conference on Artificial Intelligence*, September 2015.

[20] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing.* Birkhäuser, 2013.

[21] F. Graf, H.-P. Kriegel, M. Schubert, S. Pölsterl, and A. Cavallaro. 2D Image Registration in CT Images using Radial Image Descriptors. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, pages 607–614. Springer, September 2011.

[22] F. Graf, H.-P. Kriegel, M. Schubert, S. Pölsterl, and A. Cavallaro. Position prediction in ct volume scans. In *Proceedings of the 28th International Conference on Machine Learning (ICML) Workshop on Learning for Global Challenges*, 2011.

[23] David Harrison and Daniel L. Rubinfeld. Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Management*, 5(1):81–102, 1978.

[24] Laurent Jacques, David K. Hammond, and Jalal M. Fadili. Dequantizing Compressed Sensing: When Oversampling and Non-Gaussian Constraints Combine. *IEEE Transactions on Information Theory*, 57(1):559–571, January 2011.

[25] Ron Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.

[26] M. Lichman. UCI Machine Learning Repository, 2013.

[27] Philipos C. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press, 2013.

[28] Dirk A. Lorenz, Marc E. Pfetsch, and Andreas M. Tillmann. Solving Basis Pursuit: Heuristic Optimality Check and Solver Comparison. *ACM Transactions on Mathematical Software*, 41(2):Article No. 8, January 2015.

[29] David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*. Springer, third edition, 2008.

[30] Roberto López González. *Neural Networks for Variational Problems in Engineering*. PhD thesis, Technical University of Catalonia, 2008.

[31] Julien Mairal and Bin Yu. Complexity Analysis of the Lasso Regularization Path. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 353–360, July 2012.

[32] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[33] B. K. Natarajan. Sparse Approximate SSolution to Linear Systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.

[34] J. L. Nazareth. The Homotopy Principle and Algorithms for Linear Programming. *SIAM Journal on Optimization*, 1(3):316–332, August 1991.

[35] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, second edition, 2006.

[36] M. R. Osborne, Brett Presnell, and Turlach B. A. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20(3):389–404, July 2000.

[37] Haotian Pang, Tuo Zhao, Robert Vanderbei, and Han Liu. A Parametric Simplex Approach to Statistical Learning Problems. http://www.princeton.edu/∼rvdb/tex/PSM/PSM.pdf, 2015.

[38] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1972.

[39] David Salomon. *Data Compression: The Complete Reference*. Springer Science & Business Media, 2004.

[40] Alexander Schrijver. *A Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.

[41] Kishan Shenoi. *Digital Signal Processing in Telecommunications*. Prentice Hall PTR, 1995.

[42] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[43] Joel A. Tropp. Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[44] Robert J. Vanderbei. *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, 2nd edition, 2001.

[45] Peter Vary and Rainer Martin. *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. John Wiley & Sons, 2006.

[46] Songfeng Zheng and Weixiang Liu. An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Computers in Biology and Medicine*, 41(11):1033–1040, November 2011.